



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Ph.D. Dissertation of Spanish Linguistics

**Linguistic alignment classification
and generation with deep learning
in Spanish conversation**

스페인어 대화문에서의
딥러닝 기반 인접 호응 분류 및 생성 연구

February 2018

Graduate School of Humanities
Seoul National University
Spanish Linguistics Major
Eunchung Noh

Linguistic alignment classification and generation with deep learning in Spanish conversation

Un-Kyung Kim

Submitting a Ph.D. Dissertation of Spanish Linguistics

December 2017

Graduate School of Humanities

Seoul National University

Spanish Linguistics Major

Eunchung Noh

Confirming the Ph.D. Dissertation written by

Eunchung Noh

January 2018

Chair	<u>Hyo Pil Shin</u>	(Seal)
Vice Chair	<u>Un-Kyung Kim</u>	(Seal)
Examiner	<u>Min-Hwa Chung</u>	(Seal)
Examiner	<u>Do-Hyung Kang</u>	(Seal)
Examiner	<u>Sang-Young Mah</u>	(Seal)

Abstract

Linguistic alignment classification and generation with deep learning in Spanish conversation

Eunchung Noh

Spanish Linguistics Major

Graduate School of Humanities

Seoul National University

In general communication, the speaker changes the style of speech in order to accept or support the opinion of the other person, and changes the point of communication according to the previous interlocutor's speech. Here we constructed a tagged corpus for alignment, as well as a model for predicting the linguistic alignment markers. In this study, the alignment refers to the reaction or attitude of recognizing previous speech and the strategy of choosing words according to the context. This phenomenon is observed in many different contexts and may vary depending on various social factors and causes. It is a kind of accommodation or adaptation behavior, commonly observed in web-based conversations as well as face-to-face situations. Therefore, it can be used to infer the social dynamics that are potentially difficult to directly observe the immediate response. Until now, however, there was no standard method for measuring the aligned response. Its mechanism is unclear

whether it is based on the structural level, or the purpose of the conversation. In order to answer these questions, we try to use the Deep Learning Prediction Model to find the alignment structure.

Although alignment structure is often found in real world, there was no clear standard in previous studies, we started to define it theoretically. Alignment response refers to a structure that continues conversation in line with the context of the immediately preceding utterance. It affirms or rejects the utterance of the other party, and repeats the preceding statement. On the other hand, non-alignment refers to cases in which the above-mentioned topics are changed or the flow of the conversation is interrupted, and the words are not related to each other at all. Linguistic alignment is observed in many situations and can vary depending on various social factors and causes.

It is difficult to analyze the alignment response structure in a conventional rule-based method in that it is a problem that is obviously existent but is difficult to define in practice. Therefore, we tried to analyze and utilize the state-of-the-art deep learning method as a solution.

This study has three novelties. Firstly, we applied deep learning for the first time to detect and classify linguistic alignment. Secondly, we applied the concept of alignment response to Spanish for the first time. We design RNN, CNN, and Siamese network models to propose a method for establishing theoretical criteria of alignment sentences in Spanish quasi-spoken corpus. Thirdly, we validated the Spanish vocabulary markers.

Spanish has a very well developed vocabulary structure with lots of explicit lexical elements. We analyzed the vocabulary and category characteristics with four textbooks of Seoul National University and crawled web Twitter data. We created a crawler for collecting Twitter conversations, and constructed the correct answer data

by establishing the linguistic theoretical criteria in which the alignment responded in the collected corpus. For a model that does not depend on a specific marker, a deep learning method that takes into account the whole sentence is applied. We developed a classifier for predicting the alignment using the RNN, CNN, and Siamese networks. Experiments have shown that it is difficult to classify Spanish sentences based on specific vocabulary tokens.

In order to train alignment classification, we analyzed the twitter conversation and the dialogues in Spanish textbooks of Seoul National University. Based on the established response criteria, we classify 2,000 pairs of twitters and 1,384 dialogue pairs in textbooks, and validated the quality of training using the valid loss model selection method. The classifier was trained by 8,400 pairs of augmentation of each data at 1: 1 ratio (1,960 pairs of alignment, 1,960 pairs of non-alignment in L2 acquisition materials; 2,250 pairs of alignment and 2,250 pairs of non-alignment in Twitter conversion). The performance of the classifier was fairly good (70-80 %), and higher than 35-55 % of the baseline (classified as alignment response if the markers existed in the answer sentence) based on the theory.

The unlabeled data of 50,000 pairs (alignment 27,449 pairs, non-alignment 22,551 pairs) were classified through the trained classifier, and we made a generation model using 27,449 pairs of alignment-classified dialogues. As a result, we could generate various responses with better quality from the classified data than those from labeled data. This shows that using the trained classifier is a useful way to collect a large amount of data and reduce human labor. We also validated markers based on classified Spanish data, and found unexpected markers. We confirmed that deep learning can be very useful in linguistic research, and it is possible to design a system that generates a context-sensitive response, depending on the user 's input.

This is the first study to establish the theoretical concept of the linguistic align-

ment and to validate the markers by deep learning to solve this difficult problem. We have tried to verify the markers by detecting the aligned response syntax and modeling them through various experiments, and achieved the desired purpose of generating an answer to the alignment response. The problem we mentioned is very important in practice, and the lack of high performance of the model reflects the difficulty of the task itself. We are expecting this to be a basic deep learning methodology for linguistic research and developing a system that will solve the data shortage and respond appropriately to users.

Keywords : Linguistic Alignment, Spanish Marker, Alignment Classification, Marker Validation, Aligned Response Generation, Quasi-spoken Corpus

Student Number : 2008-30759

Table of Contents

Abstract	i
1. Introduction	1
2. Background	5
2.1 Linguistic Alignment	5
2.1.1 Linguistic Alignment in short dialogues	7
2.1.2 Linguistic Alignment in long dialogues	8
2.1.3 Definition and Criteria	9
2.1.4 Related Work	16
2.2 Natural Language Processing with Deep Learning	19
2.2.1 Traditional Natural Language Processing (NLP)	19
2.2.2 Deep Learning Approach	21
2.2.3 Recurrent Neural Network (RNN)	30
2.2.4 Convolutional Neural Network (CNN)	33
2.2.5 Siamese Network	37
2.2.6 Sequence to Sequence (Seq2Seq)	38
2.2.7 Attention Model	38
2.3 Chapter Summary	45
3. Experiment	47
3.1 Overview	47
3.2 Data Acquisition	51
3.2.1 Data Types	51

3.2.2	Crawling and Preprocessing	52
3.2.3	Data Labeling	53
3.2.4	Data Augmentation	54
3.3	Alignment Classification	58
3.3.1	Baseline - marker-based classification	58
3.3.2	Settings	60
3.3.3	Results	79
3.3.4	Qualitative Evaluation	86
3.4	Generation trained with labeled pairs	88
3.4.1	Settings	88
3.4.2	Results	89
3.4.3	Qualitative Evaluation	89
3.5	Generation trained with classified pairs	95
3.5.1	Settings	97
3.5.2	Results	97
3.5.3	Qualitative Evaluation	97
3.6	Marker Validation	103
3.6.1	Marker extraction from deep learning model	103
3.6.2	Validation of Marker from human's and model's	104
3.6.3	Qualitative Evaluation	105
3.7	Chapter summary	115
4.	Discussion	117
4.1	Linguistic Feature of Spanish	117
4.2	Advantage of Deep Learning	123
5.	Conclusion	125

References	127
Appendix	137
A.1 Crawling and Preprocessing - Whole codes	137
A.2 Alignment Classifier - Whole codes	145
A.3 Generation of Aligned-conversation - Whole codes	166
A.4 Marker Validation - Whole codes	177
A.5 Spanish Conversation Pairs	180
A.5.1 Textbook corpus	180
A.5.2 Twitter corpus	215

List of Figures

Fig. 1.	Alignment of a single pair of dialogues	7
Fig. 2.	Alignment of a dialogue with several turns	8
Fig. 3.	Definition of alignment	10
Fig. 4.	Flowcharts of the different AI systems.	22
Fig. 5.	An illustration of Stochastic gradient descent (SGD)	25
Fig. 6.	An illustration of dropout	26
Fig. 7.	An illustration of skip-gram	27
Fig. 8.	An illustration of word embedding	28
Fig. 9.	An illustration of copy mechanism for machine translation	29
Fig. 10.	Long short term memory and Gated Recurrent Unit	32
Fig. 11.	Convolutional neural network for text processing.	35
Fig. 12.	Model architecture with two channels for an example sentence.	37
Fig. 13.	Siamese LSTM	38
Fig. 14.	Siamese CNN	39
Fig. 15.	Encoder-decoder model	40
Fig. 16.	An illustration of attention mechanism	41
Fig. 17.	Example of attention matrix	42
Fig. 18.	An illustration of attention over attention	43
Fig. 19.	An illustration of self-attention	44
Fig. 20.	Alignment classification with deep learning classifier and marker-based approach	48
Fig. 21.	Aligned response generation with NMT model	49

Fig. 22.	Selection of alignment in textbook (<i>Español Básico</i> 1, Chpt 10, p.172)	54
Fig. 23.	Selection of alignment in Twitter	55
Fig. 24.	An illustration of cross-validation	56
Fig. 25.	An illustration of early stopping	57
Fig. 26.	Models used in the experiment	77
Fig. 27.	Best results of models	80
Fig. 28.	BiLSTM result with textbook data.	82
Fig. 29.	CNN result with textbook data.	82
Fig. 30.	Siamese BiLSTM result with textbook data.	83
Fig. 31.	Siamese CNN result with textbook data.	83
Fig. 32.	BiLSTM result with twitter data.	84
Fig. 33.	CNN result with twitter data.	84
Fig. 34.	Siamese BiLSTM result with twitter data.	85
Fig. 35.	Siamese CNN result with twitter data.	85
Fig. 36.	Attention matrix from Table 39 1st example	93
Fig. 37.	Attention matrix from Table 39 2nd example	94
Fig. 38.	Outline of the generation trained with classified pairs	96
Fig. 39.	Attention matrix from Table 42 1st example	101
Fig. 40.	Attention matrix from Table 42 2nd example	102
Fig. 41.	Outline of the marker validation	104

List of Tables

Table 1. Examples of disruption in short sentences	6
Table 2. Examples of disruption in long sentences	6
Table 3. Examples of disruption in dialogue	6
Table 4. Comparison of alignment definition in previous studies	9
Table 5. Alignment classification 1: Affirmation	12
Table 6. Alignment classification 2: Negation	13
Table 7. Alignment classification 3: Repetition	13
Table 8. Non-alignment classification: Switch, gap, irrelevant speech . . .	14
Table 9. Strict classification criteria for continuous textbook conversation	15
Table 10. Contrasts between autonomous transmission account of language processing in dialogue and the interactive alignment account (Pick- ering and Garrod, 2004b)	18
Table 11. Advantages and challenges of deep learning for natural language processing. (Table from Li 2017)	23
Table 12. Data composition before adjustment	52
Table 13. Examples of data augmentation using word dropout	56
Table 14. List of 20 components for verification experiments consisting of 4 levels	59
Table 15. Results of baseline experiment	60
Table 16. Alignment for directly related interpersonal conversations	61
Table 17. Examples of alignment-labeled data: Textbook (1)	63
Table 18. Examples of alignment-labeled data: Textbook (2)	64
Table 19. Examples of alignment-labeled data: Twitter (1)	65

Table 20.Examples of alignment-labeled data: Twitter (2)	66
Table 21.Examples of alignment-labeled data: Twitter (3)	67
Table 22.Examples of alignment-labeled data: Twitter (4)	68
Table 23.Examples of non-alignment data: Textbook (1)	69
Table 24.Examples of non-alignment data: Textbook (2)	70
Table 25.Examples of non-alignment data: Twitter (1)	71
Table 26.Examples of non-alignment data: Twitter (2)	72
Table 27.Examples of non-alignment data: Twitter (3)	73
Table 28.Examples of non-alignment data: Twitter (4)	75
Table 29.Augmented alignment and non-alignment data set	76
Table 30.List of Hyperparameters-BiLSTM	77
Table 31.List of Hyperparameters-CNN	78
Table 32.List of Hyperparameters-Siamese-BiLSTM	79
Table 33.List of Hyperparameters-Siamese-CNN	79
Table 34.Best results of models	81
Table 35.Evaluation of the results from the alignment classifier	87
Table 36.Hyperparameter lists of Encoder and Attention-decoder	89
Table 37.Example of generated responses from the NMT model trained with labeled pairs	90
Table 38.Evaluation of generated responses from the NMT model trained with labeled pairs	92
Table 39.Generated responses for investigating attention matrix.	93
Table 40.Example of generated responses from the NMT model trained with classified pairs	98
Table 41.Evaluation of generated responses from the NMT model trained with classified pairs	99

Table 42.Generated responses for investigating attention matrix.	100
Table 43.Candidates for markers	105
Table 44.Markers obtained from the model	106
Table 45.Alignment marker. Explicitly distinguished markers and high fluctuation of frequency rate (total appearance / total replies)	107
Table 46.Non-alignment marker. Explicitly distinguished markers and high fluctuation of frequency rate (total appearance / total replies) . .	107
Table 47.Comparison of the eight vocabulary from alignment sentence pairs only	108
Table 48.Eight examples with alignment markers	109
Table 49.Feature specification of subject/object constructions in Spanish and English (Montrul 2004:130)	121
Table 50.The syntactic differences between Spanish and English with respect to the Null Subject Parameter (Toribio, 2000)	122
Table A1. Examples of labeled data: Textbook (1)	181
Table A2. Examples of labeled data: Textbook (2)	182
Table A3. Examples of labeled data: Textbook (3)	183
Table A4. Examples of labeled data: Textbook (4)	184
Table A5. Examples of labeled data: Textbook (5)	185
Table A6. Examples of labeled data: Textbook (6)	186
Table A7. Examples of labeled data: Textbook (7)	187
Table A8. Examples of labeled data: Textbook (8)	188
Table A9. Examples of labeled data: Textbook (9)	189
Table A10.Examples of labeled data: Textbook (10)	190
Table A11.Examples of labeled data: Textbook (11)	191
Table A12.Examples of labeled data: Textbook (12)	192

Table A13.Examples of labeled data: Textbook (13)	193
Table A14.Examples of labeled data: Textbook (14)	194
Table A15.Examples of labeled data: Textbook (15)	195
Table A16.Examples of labeled data: Textbook (16)	196
Table A17.Examples of labeled data: Textbook (17)	197
Table A18.Examples of labeled data: Textbook (18)	198
Table A19.Examples of labeled data: Textbook (19)	199
Table A20.Examples of labeled data: Textbook (20)	200
Table A21.Examples of labeled data: Textbook (21)	201
Table A22.Examples of labeled data: Textbook (22)	202
Table A23.Examples of labeled data: Textbook (23)	203
Table A24.Examples of labeled data: Textbook (24)	204
Table A25.Examples of labeled data: Textbook (25)	205
Table A26.Examples of labeled data: Textbook (26)	206
Table A27.Examples of labeled data: Textbook (27)	207
Table A28.Examples of labeled data: Textbook (28)	208
Table A29.Examples of labeled data: Textbook (29)	209
Table A30.Examples of labeled data: Textbook (30)	210
Table A31.Examples of labeled data: Textbook (31)	211
Table A32.Examples of labeled data: Textbook (32)	212
Table A33.Examples of labeled data: Textbook (33)	213
Table A34.Examples of labeled data: Textbook (34)	214
Table A35.Examples of labeled data: Textbook (35)	215
Table A36.Examples of labeled data: Twitter (1)	216
Table A37.Examples of labeled data: Twitter (2)	217
Table A38.Examples of labeled data: Twitter (3)	218

Table A39.Examples of labeled data: Twitter (4)	219
Table A40.Examples of labeled data: Twitter (5)	220
Table A41.Examples of labeled data: Twitter (6)	221
Table A42.Examples of labeled data: Twitter (7)	222
Table A43.Examples of labeled data: Twitter (8)	223
Table A44.Examples of labeled data: Twitter (9)	224
Table A45.Examples of labeled data: Twitter (10)	225
Table A46.Examples of labeled data: Twitter (11)	226
Table A47.Examples of labeled data: Twitter (12)	227
Table A48.Examples of labeled data: Twitter (13)	228
Table A49.Examples of labeled data: Twitter (14)	229
Table A50.Examples of labeled data: Twitter (15)	230
Table A51.Examples of labeled data: Twitter (16)	231
Table A52.Examples of labeled data: Twitter (17)	232
Table A53.Examples of labeled data: Twitter (18)	233
Table A54.Examples of labeled data: Twitter (19)	234
Table A55.Examples of labeled data: Twitter (20)	235
Table A56.Examples of labeled data: Twitter (21)	236
Table A57.Examples of labeled data: Twitter (22)	237
Table A58.Examples of labeled data: Twitter (23)	238
Table A59.Examples of labeled data: Twitter (24)	239
Table A60.Examples of labeled data: Twitter (25)	240
Table A61.Examples of labeled data: Twitter (26)	241
Table A62.Examples of labeled data: Twitter (27)	242
Table A63.Examples of labeled data: Twitter (28)	243
Table A64.Examples of labeled data: Twitter (29)	244

Table A65.Examples of labeled data: Twitter (30)	245
Table A66.Examples of labeled data: Twitter (31)	246
Table A67.Examples of labeled data: Twitter (32)	247
Table A68.Examples of labeled data: Twitter (33)	248
Table A69.Examples of labeled data: Twitter (34)	249
Table A70.Examples of labeled data: Twitter (35)	250
Table A71.Examples of labeled data: Twitter (36)	251
Table A72.Examples of labeled data: Twitter (37)	252
Table A73.Examples of labeled data: Twitter (38)	253
Table A74.Examples of labeled data: Twitter (39)	254
Table A75.Examples of labeled data: Twitter (40)	255
Table A76.Examples of labeled data: Twitter (41)	256
Table A77.Examples of labeled data: Twitter (42)	257
Table A78.Examples of labeled data: Twitter (43)	258
Table A79.Examples of labeled data: Twitter (44)	259
Table A80.Examples of labeled data: Twitter (45)	260
Table A81.Examples of labeled data: Twitter (46)	261
Table A82.Examples of labeled data: Twitter (47)	262
Table A83.Examples of labeled data: Twitter (48)	263
Table A84.Examples of labeled data: Twitter (49)	264
Table A85.Examples of labeled data: Twitter (50)	265

List of Codes

Code 1. Crawler.py	137
Code 2. arrange_retweet.py	139
Code 3. Preprocessing code with library <i>Regular Expression (RE)</i>	141
Code 4. Tokenizing with library <i>Natural Language Tool Kit (NLTK)</i>	141
Code 5. Separating train data into a training set and a validation set with library <i>scikit-learn</i>	141
Code 6. split_train_valid.py	142
Code 7. arrange_data_no_label.py	143
Code 8. baseline.py	145
Code 9. augment_word_dropout_separately.py	146
Code 10. main.py	148
Code 11. utils.py	155
Code 12. models.py	161
Code 13. For hyper-parameter tuning, multi-train_text.sh	165
Code 14. main_enc_dec.py	166
Code 15. Encoder (BiLSTM)	173
Code 16. Decoder (LSTM)	174
Code 17. Attention-Decoder (LSTM with attention)	175
Code 18. find_marker.py	177

Our github URL is <https://github.com/eunchung/Alignment>, containing twitter data and whole codes for this dissertation.

Chapter 1

Introduction

In contemporary linguistic research, there is a tendency to expand the analysis of discourse on various practical texts. It can be confirmed from the fact that attempts to approach the various types of speech in the spoken language, dialogue, and discourse structure. In addition, the research method of approaching grammatical morphemes, which have been mainly discussed in syntactic theory or semantics has emerged through discourse data. The analysis data is not limited to the written data but is expanded to the spoken language. The proportion of the spoken data tends to be higher than the written language. And modern natural language processing (NLP) technology is expected to be utilized as a base technology for providing natural language interface between human and computer.

Natural language processing is a computational technique for automating human language analysis and representation. NLP research enables computers to carry out various tasks such as machine translation and dialogue systems. In particular, when a pair of sentences is given, to decide how to express and how to establish the relationship is one of the most important tasks in NLP tasks. These kinds of work include answer selection, paraphrase identification, and textual entailment in the question and answer pairs.

This study introduces the concept of alignment and applies it to Spanish conversation. Linguistic alignment can be a kind of linguistic adaptation phenomenon. Some previous studies (Pickering and Ferreira, 2008; Kaschak et al., 2011; Reitter et al., 2011; Wang et al., 2014) give theoretical foundations of linguistic adaptation

of the language patterns to match conversational partners, by using word choice, sentence structure, and so on. Especially, in a conversational pair, the phenomenon that the following speaker's answer is affected by the previous speech in order to respond to the immediate utterance of the other is called '*alignment*' (Noh, 2017a,b). Experimental results provide a clearer understanding of the linguistic alignment of the Spanish language belonging to the Romance group where many features of Latin remain, and can be used to create a dialogue system that generates aligned responses based on user's speech.

Until now, there was no standard for defining or measuring the aligned response. Its mechanism is still unclear. Linguistic alignment is observed in many situations and can vary depending on various factors and causes. Therefore, it is difficult to define alignment responses in practice and to analyze its patterns in a conventional rule-based method. In order to measure and validate the markers, we use the state-of-the-art deep learning method and detect the alignment pairs from the prediction model.

Purpose of the study In the two - person dialogue, we try to analyze and model the user's utterance intentions revealed through narratives and emotions. We want to create a model that predicts the alignment sentence pairs.

Until now, there was no standard method for measuring the aligned response by preliminary studies in which the word category or lexicon was added together, and the vocabulary, the category, and the concept were unclear. It is necessary to focus on the vocabulary and the categorical characteristics in a language in which formal markers such as pronouns and interrogative expressions are explicitly revealed. This is the very first study to validate markers based on prediction models using deep learning.

Spanish is the second most spoken language in the world, and the number of users is still increasing. As the number of Internet users increases, the amount of data in Spanish is also exploding. Developing a system for Spanish speakers seems beneficial not only in academic research but also in practical usage.

The contributions of this paper are:

1. we applied deep learning for the first time to detect and classify linguistic alignment.
2. we applied the concept of alignment response to Spanish for the first time. We design RNN, CNN, and Siamese network models to propose a method for establishing theoretical criteria of alignment sentences in Spanish quasi-spoken corpus.
3. we constructed a tagged corpus for alignment, as well as a model for predicting the linguistic alignment markers.

Outline of the thesis The composition of this paper is as follows. In chapter 2, we introduce the concept of linguistic alignment and the deep learning method used in the model to classify alignments. In chapter 3, alignment classification, aligned response generation, and marker validation are described. In chapter 4, we discussed about linguistic feature of Spanish and advantage of deep learning. Lastly we conclude in chapter 5.

Chapter 2

Background

2.1 Linguistic Alignment

People tend to agree or disagree with the opinion and follow the interlocutor's style of speech in communication. There are several ways to follow the speech style, such as making their sentence structure similar and using some vocabulary repeatedly. Besides these methods, people may adapt or imitate each other in pronunciation, posture, attitude, voice or speed, to unconsciously feel a sense of identity to their conversational partners Doyle and Frank (2016); Doyle et al. (2017); Danescu-Niculescu-Mizil et al. (2013).

A sentence pairs are properly connected by the words mentioned in the preceding and following sentences. In a coherent structure, the speaker's utterances contain necessary information and do not lack explicit or meaningful information. In Table 1, however, when coherence is broken within a sentence, the sentence becomes contradictory.

The longer the sentence length is, the more the coherence is broken. Sentences without cohesion often contain meaningless contents, as in Table 2. The tense agreement between sentences may also be broken or scrambled. And this kind of disruption may be observed as well in dialogue with other people, as in Table 3.

People tend to use similar words to their partners in a conversation, as a kind of acceptance behavior, showing that they adapt to the behavior or words of others during the conversation. This situation is observed in computers and web-based con-

Table 1: Examples of disruption in short sentences

Spanish	English
Pepe es un cobarde valiente. Sus ojos ciegos les miraban fijamente. Las arenas mojadas estaban más secas que nunca. Mi mamá tiene menos edad que yo. Mi color favorito es el naranja pero también odio ese color. Juana me acompañó pero no me acompañó.	Pepe is a brave coward. Their blind eyes stared at them. The wet sands were drier than ever. My mom is younger than me. My favorite color is orange but I also hate that color. Juana accompanied me but did not accompany me.

Table 2: Examples of disruption in long sentences

Spanish	English
Mi rata, que era de color verde, salió por la galaxia. Dio una vuelta y trajo un cometa pero yo ya había hecho las compras del día y entonces llegó el campeonato de básquet y perdimos.	My rat, which was green, came out through the galaxy. He came around and brought a kite but I had already made the purchases of the day and then the basketball champion arrived and we lost.

Table 3: Examples of disruption in dialogue

Speaker	Spanish	English
A	¿Fuiste a correr ayer?	Did you go running yesterday?
B	¡Sí!	Yes!
A	¿Y cuánto has corrido?	And how much have you run?
B	No, yo no he comido nada.	No, I haven't eaten at all.

versations as well as face-to-face situations. And this phenomenon can affect not only the social factors such as the influence of the other party and the social network centrality, but also personal liking. Therefore, there is a possibility that it can be used to infer the social dynamics that are potentially difficult to directly observe the immediate response. However, it is not well understood whether it is based on the underlying mechanism, its qualities, the structural level, or the purpose of the conversation. In order to answer these questions, we try to find the appropriate markers by using the

deep learning prediction model for the aligned sentences. Some word tokens may do special function in conversation and work as alignment markers in Spanish.

In this study, we investigate the linguistic alignment in the following two cases.

2.1.1 Linguistic Alignment in short dialogues

In a short dialogue consisting of one turn of Speaker 1 and Speaker 2, the adjoining response consists of repeating the word directly or answering the preceding utterance.

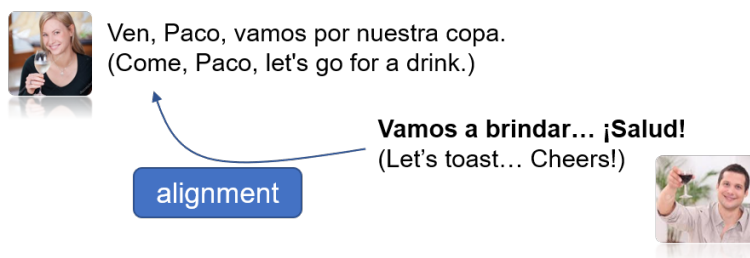


Fig. 1: Alignment of a single pair of dialogues

In the example Fig. 1, the second speaker repeatedly used the word "vamos" as spoken by the first speaker. In practice, however, the "copa" mentioned by the first speaker conveys the main meaning, and the second speaker has thus chosen words such as "brindar" and "salud". It is not enough to define as changing word selection or comparing similarity to the previous sentence. Therefore, it is necessary to redefine the concept in terms of recognizing and reacting to previous speech and choosing vocabulary according to context (in accordance with the previous speech and the strategy of choosing words according to the context).

This study is the very first study of linguistic alignment in Spanish, setting the simplest task. In short dialogues, we define the problem of identifying alignment

from the previous sentence to the following sentence. Conversation data of speaker A / speaker B in the textbooks and tweets / replies of Twitter were judged to be available for such a relationship, in the Spanish quasi-spoken corpus.

2.1.2 Linguistic Alignment in long dialogues

A long conversation consisting of several turns, in which Speaker 1 and Speaker 2 alternately talk, is likely to be more fragile than a short conversation. The longer the dialogue, the weaker the response structure, while the relationship between the speakers can be inferred or the centrality of dialogue is also available.

In Fig. 2 the speaker B repeats what he or she has said (Nos vemos mañana por la tarde) or partly as it is la original. However, in A', there are no elements related to the contents of the previous conversation or objects to be repeated. The longer the conversation takes and the more turns, the more likely this trend will be.

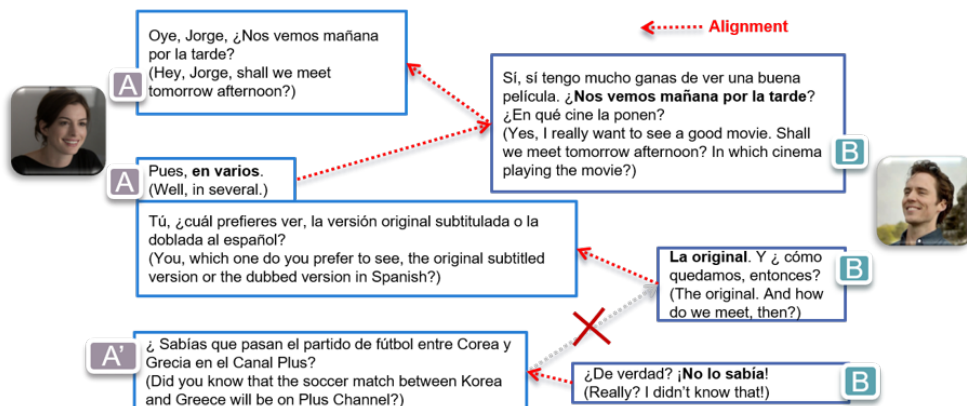


Fig. 2: Alignment of a dialogue with several turns

Table 4: Comparison of alignment definition in previous studies

Author(s)	Year	Data & Tools	Objectives	Definition
Doyle, Goldberg, Srivastava & Frank (ACL; cited 5)	2017	7-year corpus of corporate emails (alignment in the first six months)	measure of cultural fit based on linguistic alignment	Linguistic alignment: the tendency to use the same or similar words as one's conversational partner
Doyle, Yurovsky, Frank (WWW; cited 10)	2016b	Twitter, probabilistic Hierarchical Alignment Model	measures for linguistic alignment	Linguistic alignment: the amount that one person's word use is influenced by another's
Doyle, Frank (ACL; cited 4)	2016a	Switchboard corpus, Twitter, probabilistic Hierarchical Alignment Model & Word-based Hierarchical Alignment Model	sources of linguistic alignment	Linguistic alignment: change of word choice based on others' choices
Xu, Yang & Reitter (CMCL; cited 7)	2015	indiscriminate local linguistic alignment (LLA), Spearman's correlation coefficient (SCC), and repetition decay (RepDecay)	evaluation for the properties of measures	Linguistic alignment: the convergence of linguistic choices among interlocutors
Menenti (cited 46)	2012	neuroscience literature review	neural mechanisms of alignment (neural activity and its relation to mental states)	Alignment: speaking and listening make use of similar representations, and have largely overlapping neural correlates

2.1.3 Definition and Criteria

Definition and measures of alignment have varied from study to study and field to field, leading to incommensurable results. The key reference of this paper is in Table 4.

Alignment increases in the probability of seeing a given marker (or marker category) in the second message of a pair given that it appeared in the preceding message (Doyle et al., 2016). Alignment occurs at levels of structure related to meaning, which is the choice of one alternative over other associated with different semantic representations. Previous studies have focused on word selection or similarity with

previous sentences.

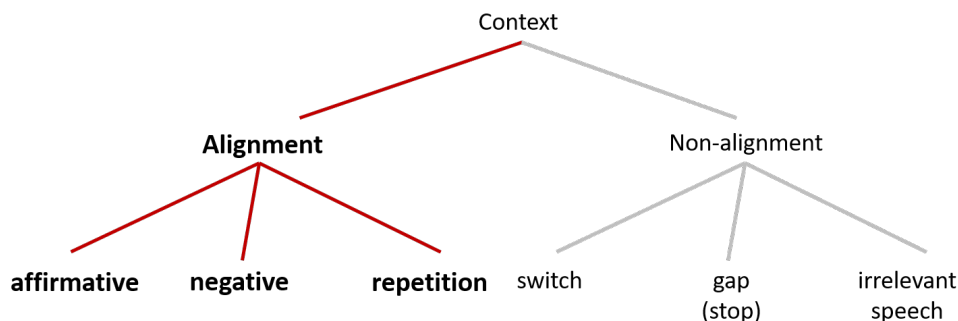


Fig. 3: Definition of alignment

In this study, the criteria of alignment are divided into three categories, as shown in 3: affirmative, negative, and repetition. Since there were no gold standards or methodologies in previous studies, we set up empirical standards that fit our research goals. And, in order to grasp the linguistic characteristics that can be overlooked at the level of the lexical or syntactic level, we chose to input the sentences of each dialogue turn.

Criteria Despite the fact that Spanish is a commonly spoken language, there are not many studies on this field. The topic of linguistic alignment is also not covered. For this purpose, we classify the alignment and nonalignment statements in the paper and set the criteria respectively.

In a two-person dialogue, if the conversation of the next person corresponds to the conversation of the previous person, classify it as an *alignment*. Based on the previous conversation, tag the next person in the case of positive, negative, and repetitive answers in alignment.

1. Select and tag the sentences that correspond to immediate conversation, including all the situations within the context.

2. If they do not correspond, classify them as non-alignment and sort them by type: switch, gap (stop), irrelevant speech.

It is important to note that not only the affirmation of the speaker but also the negation within the context are judged as *alignment*. And it is also called alignment to imitate the entire sentence or phrase of the interlocutor. On the other hand, the alignment relation can be completely changed or interrupted by irrelevant contents. If there is a completely unrelated element or a topic which is out of context, then alignment is broken (i.e., non-alignment).

Repeating certain phrases produces some extra effect such as emphasis. Spanish speakers rarely repeat without changing other words or pronouns. They usually think that abundant expressions are beauty and virtue of language. So there are many examples. It is colloquial, and usually assumes a certain situation. And the interactive alignment of dialogue provides with the central role of imitation within psychological and neuroscientific theory (Pickering and Garrod, 2004b). Mirror neurons provide a reason to expect certain forms of imitation to be straightforward, and the findings that the same areas of the brain (Brodmann's Areas 44 and 45; inferior frontal gyrus) are involved in verbal imitation support the assumption that alignment constitutes a fundamental aspect of language use.

In the tagged data, we tried to allot each alignment structures by 1/3 of total alignment data. However, there was a difference in the actual frequency. For example, in Spanish, it is rarely used repeatedly without replacing with other words or pronouns. Therefore, the proportion of repetition is relatively low compared to other alignment structures.

The non-alignment syntax was also selected for training the model and extracting the specific markers.

Each syntactic type was judged based on the used vocabulary and its meaning.

Table 5: Alignment classification 1: Affirmation

	Class	definition	
1	AFFIRMATION	Acknowledgment and acceptance about what the interlocutor has just said before	
1-1	Positive Expression	Positive vocabulary and expression in Spanish	<p>1) a. Mison: Es un viaje bastante largo. b. Casera: ¡Sí, es verdad! (Español básico 1, Chpt. 4, p.66)</p> <p>2) a. Luis: Coja a la izquierda hacia el Paseo del Prado. b. Turista: Bien, bien, entiendo. (Español básico 2, Chpt.9, p.128)</p>
1-2	Backchannels	Expressions used to reconcile the opinion of the other person, without special meaning as an individual word separately	<p>1) a. RT @MargaWirkierman: Te sorprendias con Abal Medina q estaba Duro Duro Duro?? Mira la Gobernadora en Campaña ! b. @MargaWirkierman @MechiAfonso Mmmmmmm que pasó?</p> <p>2) a. Mujer: Oh... me alegro. b. Guionista: Ah... Te alegras. (Morir)</p> <p>3) a. RT @ziberia1: Casi me echan del bar de la risa b. @ziberia1 Jajajajaja espectacular @mmmahle</p>
1-3	Semantic relevance	Recognition and semantic connectedness even if direct expression is not explicitly revealed	<p>1) a. ¿Bailando bien? b. Soy ecuatoriana y nací bailando... (Español básico 2, Chat.7, p.100)</p> <p>2) a: Lo quiero comprar. b: Te lo puedo dejar. (Español básico 1, Chpt. 9, p.166)</p>
		No similarities in form	<p>1) a. Pedro: Acabamos de ganar el partido de fútbol. b. Carla: ¡Enhorabuena! (Español básico 1, Chpt. 10, p.172)</p> <p>2) a. María: Encantada de conocerte. b. Pedro: Igualmente. (Español básico 1, Chpt. 1, p.12)</p>
1-4	Idiomatic expression	Used as social conventions	<p>1) a. Carlos: ¿Qué tal? b. Maria: ¡Hola! (Español básico 1, Chpt. 6, p.98)</p> <p>2) a. Sofia: ¡Mucho gusto! b. Quique: Encantado. (Español básico 1, Chpt. 3, p.46)</p> <p>3) a. Mison: Mil gracias. b. Casera: ¡De nada! (Español básico 1, Chpt. 4, p.66)</p>

Table 6: Alignment classification 2: Negation

	class	definition	
2	NEGATION	Refusal of the previous interlocutor's statement or oppose the opinion	
2-1	Negative Expression	Negative vocabulary and expression in Spanish	1) a. ¿Está soltero? b. No, está casado. (<i>Español básico</i> 1, Chpt. 4, p.70) 2) a. Pilar: Perdón por molestarte con mis problemas. b. Nerea: Ni hablar. No te preocupes. ¿Para qué con las amigas? (<i>Español intermedio</i> , Chat.10, p.110)
2-2	Semantic relevance	Contrary to the previous speaker's words	1) a. ¿Te importaría abrir la ventana? b. ¡Claro que no! (<i>Español básico</i> 2, Chpt.8, p.118) 2) a. Esa chica de quien me hablas fue mi primer amor. b. ¡No me digas! No lo sabía. (<i>Español intermedio</i> , Chpt.12, p.130)
		Sarcasm	(Cuando sucede algo malo) ¡Genial! ¡Estupendo! ¡Eso es lo que necesitamos!

Table 7: Alignment classification 3: Repetition

	class	definition	
3	REPETITION	Speak again the same word, phrase, or sentence the interlocutor said	
3-1	Greeting	In the case of greeting, the speaker often repeats the same sentence.	1) a. Luis: Buenos días. b. Entrevistadora: Buenos días. (<i>Español básico</i> 1, Chpt. 3, p.46) 2) a. Carlos: ¡Hola! b. Sumi: ¡Hola! (<i>Español básico</i> 1, Chpt.2, p.26)
3-2	Interrogative	To emphasize or confirm if the contents are correct	1) a. AMANTE: Entonces... ¿nos vamos? b. CORA: Nos vamos. (<i>Prohibido Suicidarse en Primavera</i>) 2) RT @PirryTv: A veces creo que todas estas traiciones entre políticos no son mas que puro teatro ... >> @PirryTv A veces??? Jajajaja. Siempre!
3-3	Imitation	To follow what someone said, without replacing with another expression	1) a. DOCTOR: Esperemos. b. HANS: Esperemos. 2) a. FERNANDO: Desde lo alto de la galería. b. AMANTE: Eso, así: desde lo alto. (<i>Prohibido Suicidarse en Primavera</i>)

Table 8: Non-alignment classification: Switch, gap, irrelevant speech

	class	definition	
1	SWITCH	Transition to other topic, interrupting in the middle, changing contents of the conversation	e.g., RT @nathcampost: Yo ayer después de ver Coco >> @nathcampost En qué tienda comercial estuvistes?
2	GAP (STOP)	Retweet without saying anything	≡ Repetition, which repeats exactly the same previous output e.g., RT @LadyGhostVzla: Siempre se hacen bromas de la mala memoria del venezolano, pero jamás pensé que muchos se olvidarían de la lucha de ello... >> @LadyGhostVzla @TITORODRIGUEZZ @LadyBloodV @AnonymousAndino @LaraAnons @NaranjaRA @CheoRebelde @txapulincolorid... https://t.co/vo56QQMFLP
		End of the conversation: Silence. No reaction to previous speech.	The subsequent conversation can not be linked to the preceding sentences. a. MANTE: ¡Cora!... ¡Cora!... b. (Silencio.) (<i>Prohibido Suicidarse en Primavera</i>)
3	IRRELEVANT SPEECH	Contents and expression completely independent of the flow of conversation.	Not related to the previous context. a. PROFESOR: Por ejemplo: qué bien está usted con ese collar de esmeraldas. b. SANTILLA: ¡Protesto en nombre de los países pobres y calientes! (<i>Siete gritos en el mar</i>)

The type of Gap is the termination of the conversation and often appears on Twitter. It may also occur in the real-world situation, however, it is beyond our research topic. We aim at generating aligned response model, and it is meaningless to train the intentionally unanswered sentence.

Table 9: Strict classification criteria for continuous textbook conversation

Dialogue 1		
¿Qué tal, María? (En. How are you Maria?)	¡Hola, Carlos! (En. Hi, Carlos!)	Alignment
¡Hola, Carlos! (En. Hi, Carlos!)	¿Estás ocupada este fin de semana? (En. Are you busy this weekend?)	Non-alignment (switch)
¿Estás ocupada este fin de semana? (En. Are you busy this weekend?)	Sí, mañana es el cumpleaños de mi abuelo. (En. Yes, tomorrow is my grandfather's birthday.)	Alignment
Dialogue 2		
¿Dónde vives, Mison? (En. Where do you live, Mison?)	Vivo en el centro de la ciudad en una pensión familiar. (En. I live in the center of the city in a family pension.)	Alignment
Vivo en el centro de la ciudad en una pensión familiar. (En. I live in the center of the city in a family pension.)	¡Qué bien hablas español! (En. How well you speak Spanish!)	Non-alignment (irrelevant)
¡Qué bien hablas español! (En. How well you speak Spanish!)	Gracias, pero no tanto. (En. Thanks, but not so much.)	Alignment
Dialogue 3		
¿Cuál es su profesión? (En. What is your profession?)	Soy ingeniero electrónico. (En. I'm an electronic engineer.)	Alignment
Soy ingeniero electrónico. (En. I'm an electronic engineer.)	¿Cuántos idiomas habla? (En. How many languages do you speak?)	Non-alignment (switch)
¿Cuántos idiomas habla? (En. How many languages do you speak?)	Hablo tres idiomas: español, francés e inglés. (En. I speak three languages: Spanish, French, and English.)	Alignment

We refined the criteria several times while excluding the subjective factors as much as possible. As the criteria become more stringent, the alignment classification and aligned response generation task becomes much harder. In the textbook data, firstly we assumed that they are almost all alignment structure, but as the non-alignment criterion became strict, some pairs are judged to be non-alignment. For more detailed description see Table 5, 6, 7, 8, and Table 9.

2.1.4 Related Work

Some studies (Reitter and Moore, 2006; Fusaroli et al., 2012) provide empirical confirmation to this research. Repetition effects are sometimes moderated in response to situational requirements or framing: they can vary in strength when the speaker (believes to) communicate with computers (Branigan et al., 2010). Repetition intensifies when the purpose of conversation is to collaborate on a common task (Reitter and Moore, 2006). Of course, communication between people is not only a linguistic but social event. Social relationships in film scripts (Danescu-Niculescu-Mizil and Lee, 2011) can be found as a cue example. A more specific aspect of language-based interaction is pragmatic convention including turn-taking, shifts in topic, and so on.

Linguistic alignment in social interaction may happen consciously. Garrod and Pickering (2009) suggest that people flexibly adapt their linguistic patterns to those of their interlocutor's. It's a sign of social coordination, or cultural fit (Doyle et al., 2017), enculturation or success of individuals and the groups to which they belong. Linguistic alignment also could happen unconsciously and be realized at different levels due to linguistic adaptation. Pickering and Garrod (2004a) show that conversations have linguistic coordination at lexical level. Branigan et al. (2000) and Gries (2005) suggest that priming effects exist at syntactic level.

Linguistic alignment has been found in written conversational text. Danescu-Niculescu-Mizil and Lee (2011) analyzes tweet conversations and confirms that linguistic alignment exists in written online social media. Backstrom et al. (2013) also show that people adjust their linguistic style and features on online. Besides, priming effects at syntactic level Gries (2005); Branigan et al. (2010). They introduced several quantitative measures for linguistic alignment phenomenon. Some methods of evaluation focus on linguistic events, such as the use of words, syntactic rules or a small set of expressions (Church, 2000; Reitter and Moore, 2006; Fusaroli et al., 2012) to

test intrinsic repetition. Linguistic feature similarity (Stenchikova and Stent, 2007; Danescu-Niculescu-Mizil and Lee, 2011) is also widely used to measure linguistic adaptation precisely.

The alignment also used in machine translation (MT) is for the bilingual framework in two language corpus, which is different from this study (Bahdanau et al., 2014). Navarro Colorado et al. (2004) uses the concept ‘alignment’ in order to explain that two different languages align with the same verb sense, the same number of arguments, the same syntactic function of each argument, and the same semantic features of each argument. Branigan et al. (2010) emphasized the importance of alignment in HCI. He pointed out that alignment strictly refers to interlocutors’ mental representations, not their actual behavior (Pickering and Garrod, 2004a; Costa et al., 2008). Two aspects to alignment in HCI reflect the two directions of communication: the computer’s contributions directed towards the user, and the user’s contributions directed towards the computer. The first aspect relates to computers that align with a human user’s contributions, and how such computers affect the user. The second aspect relates to humans’ alignment with a computer’s contributions.

Accommodation can even influence human-computer interactions, with people rating interactions with accommodating computer systems as more satisfying even when the conversant is known to be a computer (Nass and Lee, 2000; Branigan et al., 2010; Li et al., 2015a).

Table 10: Contrasts between autonomous transmission account of language processing in dialogue and the interactive alignment account (Pickering and Garrod, 2004b)

Autonomous transmission account	Interactive alignment account
1. Linkage between interlocutors Via sound alone – no direct links across other levels of representation.	1. Linkage between interlocutors Links representation across multiple levels of via “alignment channels.” Sound comes to encode words, linguistic information, and aspects of situational models.
2. Inference Internalized in the mind of speaker/listener: Speaker in terms of audience design; Listener in terms of bridging inference process.	2. Inference Externalized in the interaction between interlocutors via a basic interactive repair mechanism.
3. Routines Special case of language largely associated with idioms.	3. Routines Arise out of the application of the interactive alignment process. A high proportion of dialogue uses routines, which simplify both production and comprehension.
4. Self-monitoring Inner loop monitoring requires a special internal route from production to comprehension.	4. Self-monitoring Monitoring occurs at any level of representation that is subject to
5. Repair mechanisms Distinct repair mechanisms for self-repair and other-repair in dialogue.	5. Repair mechanisms The same basic repair mechanism for self-repair and other repair.
6. Linguistic representations Only need to account for the structure of isolated and complete sentences.	6. Linguistic representations Needed to deal with linked utterances in dialogue, including non-sentential “fragments.”

2.2 Natural Language Processing with Deep Learning

2.2.1 Traditional Natural Language Processing (NLP)

Types of NLP task Natural Language Processing (NLP) is huge field, consist of speech recognition, named entity recognition, automatic summarization, question answering, machine translation, sentiment analysis, document classification and so on. Among them, the sentence pair tasks such as recognizing textual entailment, spell correction, answer selection, and paraphrase identification are close to our linguistic alignment tasks.

Recognizing textual entailment (TE) is the task that given two sentences, classifying if one sentence entails the other sentence (positive TE) or one sentence contradicts the other sentence (negative TE), or non-textual entailment.

Spell correction task revises the wrong words with correct words, given two sentences where one sentence has wrong spell words in the sentence and the other sentence has correct words.

Answer selection task chooses the answer in a document sentence, given two sentences where one sentence is a question and the other sentence is the document which contains the answer information(word, phrase, and so on).

Paraphrase identification is the task that checks whether the one sentence has paraphrased to the other sentence or not.

Rule-based approach A method that uses multiple 'IF' statements to perform NLP tasks. It is based on many experiments, trials and errors, and experience of engineers. In other words, rule-based approach, or hand-designed program, is made through numerous trial and error. If you are confined to a specific environment, and you cope

with all situations that may arise in that environment, you can achieve very good performance.

It, however, is vulnerable to exceptions because it consists of 'IF' statements. Usually we can not address all situations. It is hard to anticipate every situation or as many situations as possible, and to write an 'IF' statement for each situation. It takes a lot of time and effort.

Traditional machine learning approach The traditional machine learning approaches to solve the Natural Language Processing (NLP) problem are well-known for logistic regression, SVM (Cortes and Vapnik, 1995), and hidden Markov Model (Rabiner, 1989). With well-extracted features, they are known to perform NLP tasks well.

However, obtaining well-extracted features usually requires many feature engineerings and that engineering is too expensive to handle many data. In other words, the NLP system based on traditional machine learning strongly depends on the hand-crafted features, and refining these features are often time-consuming.

2.2.2 Deep Learning Approach

In recent years, Deep learning approach based on neural networks have performed well in various NLP tasks. This trend has been fueled by the success of word embedding (Mikolov, 2010; Mikolov et al., 2013a,b) and deep-learning techniques (Socher et al., 2013; Graves et al., 2013b). Deep learning could be regarded as automated feature representation learning. This is the huge advantage of using deep learning. Contrary to the NLP system based on classic machine learning, deep learning does not need many feature engineering to perform well. Deep learning just learn feature representation instead of feature engineering (Fig. 4). We review papers and try to summarize the NLP task using deep learning. Instead of minutely explaining deep-learning techniques and models, we focus on recapping conceptual content and putting them together with the necessary details.

Advantage of deep learning for natural language processing. Advantages over previous methods have been identified.

1. First of all, deep learning approach in the NLP tasks shows better performance than traditional ones. In addition, there is still much room for improvement.
2. As mentioned earlier (Fig. 4), it reduces the burden for feature engineering. With this, it is even possible to end-to-end learning which outputs desired results by putting relatively raw data.
3. Other advantages are summarized in detail in the following table, in addition to challenges.

Neural Networks At the beginning, we should check what is neural networks and what brings the boom of deep learning. Neural networks means that the artificial

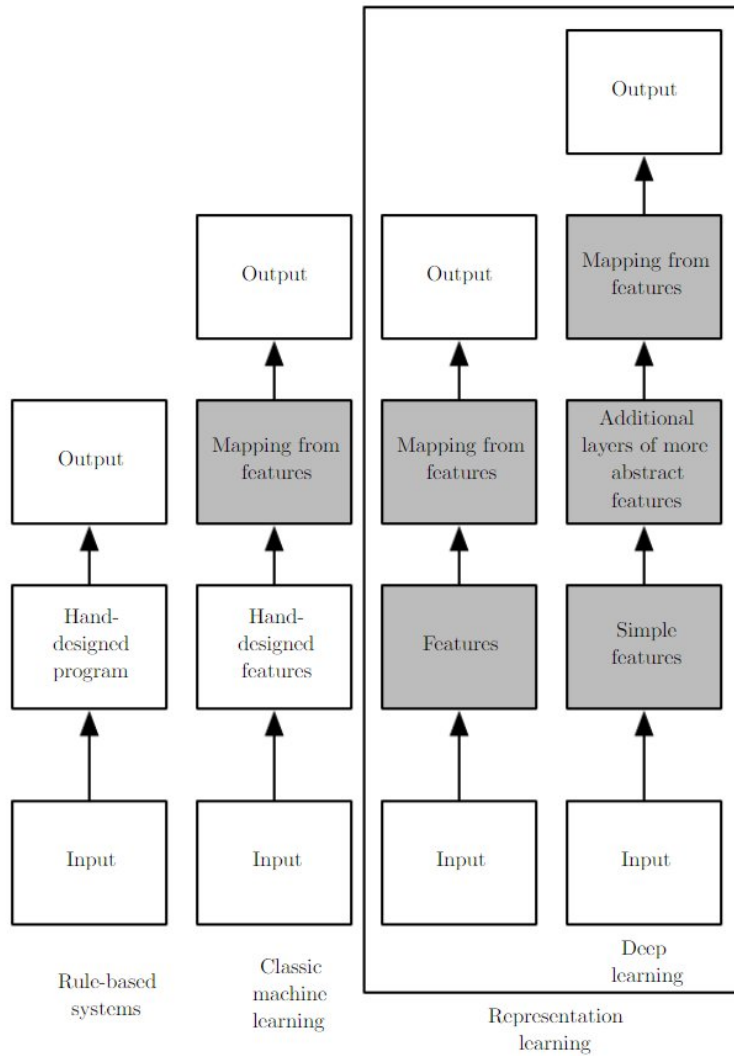


Fig. 4: Flowcharts of the different AI systems. Shaded boxes indicate components that are able to learn from data. It means that deep learning does not need hand-designed program/features, and even can do end-to-end learning (Figure from Goodfellow et al. 2016).

neural networks which mimic human brain networks. Neural networks have been a long time ago. But recently, performance has been getting better and booming. Three things are said to have contributed to the deep learning boom.

Table 11: Advantages and challenges of deep learning for natural language processing. (Table from Li 2017)

Advantages	Challenges
<ul style="list-style-type: none"> • good at pattern recognition problems • data-driven, and performance is high in many problems • end-to-end training: little or no domain knowledge is needed in system construction • learn of representations: cross-modal processing is possible • gradient-based learning: learning algorithm is simple • mainly supervised learning methods 	<ul style="list-style-type: none"> • not good at inference and decision making • cannot directly handle symbols • data-hungry and thus is not suitable when data size is small • difficult to handle long tail phenomena • model is usually a black box and is difficult to understand • computational cost of learning is high • unsupervised learning methods need to be developed • still lacks of theoretical foundation

1. A sufficient amount of data, called bigdata,
2. Graphics processing unit (GPU) calculation is introduced into the neural network,
3. Development of main algorithms such as Stochastic Gradient Descent, dropout, convolutional neural networks, recurrent neural networks and so on.

We are living in a information society. Numerous data are becoming information every day, and the number of these data exceeds trillion. It is super big data. Decades ago, there was not enough space and tools to store big data. Nowadays, each research institute and company have tools to store big data and deal with big data. That is, big data has become relatively close to us. This is the first thing that contributed to the deep learning boom.

The second is the introduction of Graphics processing unit (GPU) calculations into neural networks. As the data bigger, it takes longer to calculate. It took several

days in short, and several weeks to months in long. Raina et al. (2009) suggested that use the GPU, which usually be used for graphics operations, to the neural networks. GPU is made up of a number of cores that perform simple operations such as addition or multiplication. In other words, simple operations can be done very quickly with parallel computation. The neural network consists of a number of additions and multiplications, which is suitable for using GPU. The neural network learning through GPU reduced learning time, from several days to several hours, or from several weeks / months to several days. This is the second thing that led to the deep learning boom.

The third thing is the development of main algorithms. Typically, the development of Stochastic Gradient Descent (SGD), dropout, convolutional neural networks (CNN), recurrent neural networks (RNN) can be considered as main contribution.

There was a method to train neural network, called gradient descent (GD). Gradient descents are an effective learning method, but it takes too long time to apply large amounts of data at once. SGD is a method of learning in one data or mini batch, and it converges relatively faster than GD (Fig. 5). In other words, SGD made training of neural network faster and realistically usable. A typical SGD technique is a technique using Adam optimizer (Kingma and Ba, 2014).

Dropout is a way to train neural networks by excluding a portion of the neural net with a certain probability (Fig. 6). It causes the ensemble of several sub neural networks. As a result, the performance was noticeably improved.

Collobert et al. (2011) presented a simple deep-learning framework for NLP tasks. This framework was a state-of-the-art technique in some NLP tasks such as Named Entity Recognition (NER), Semantic Role Labeling (SRL), and POS tagging. Since then, a number of complex deep-learning-based algorithms have been proposed to solve the difficult NLP problem. There are two major deep-learning models such

²site: <https://wikidocs.net/3413>

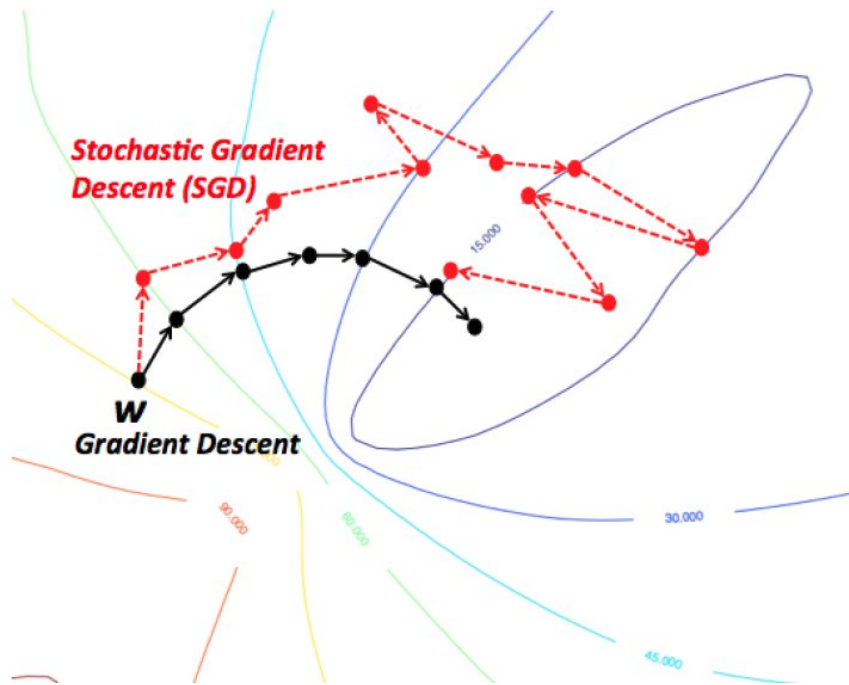


Fig. 5: Illustration of gradient descent and stochastic gradient descent (Figure from the site²)

as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). We review them in the subsequent subsection. Before talking about these two things, we cover word embedding first, which is used in almost all NLP tasks with the deep learning approach.

Word Embedding Distributional vectors or Word Embedding is the embedding of a word into a low dimensional space (eg 300 dimensions). Simply speaking, it is a technique of converting a word into a real-valued vector of 300 dimensions.

Previous machine learning methods have treated words as one-hot vectors, sparse, with higher dimensions (eg, more than 20,000 dimensions). In recent years, in various NLP assignments, we have used low-dimensional dense word embedding for superior performance.

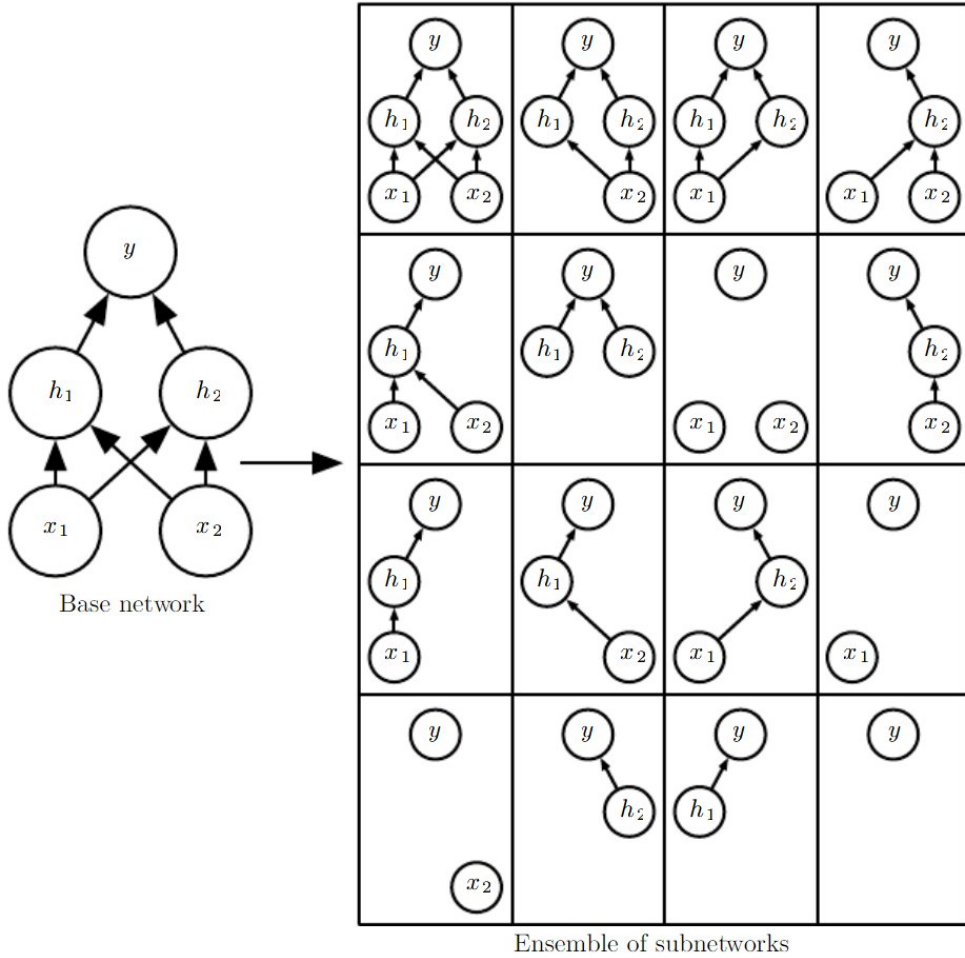


Fig. 6: Dropout trains an ensemble consisting of all subnetworks that can be constructed by removing units (except output y) from an base network. (Figure from Goodfellow et al. 2016).

Word embedding presupposes distributional hypothesis. This assumption is central to the fact that words with similar meanings will tend to appear in similar contexts. Word vectors try to capture the characteristics of neighboring words. The main advantage of word embedding is that similarity between words can be measured by using indexes such as cosine similarity between word vectors.

Word embedding is often used in the first data processing layer of the deep

learning model. In general, word embedding is pre-learned by optimizing an auxiliary objective function in an unlabeled corpus. In other words, word embedding is a byproduct of learning a certain task, such as the skip gram model (Fig. 7) that predicts the surrounding words given the current word (Mikolov et al., 2013a,b).

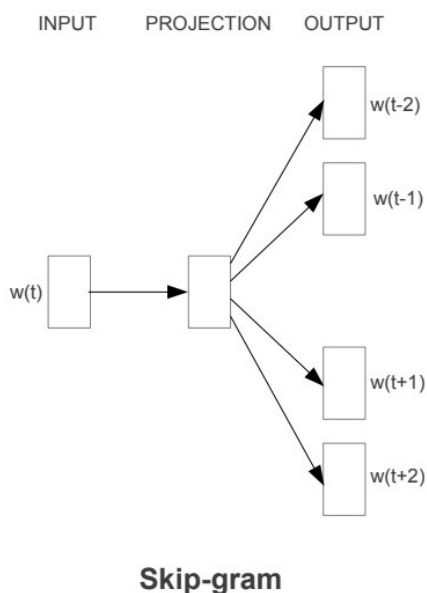


Fig. 7: An illustration of skip-gram. Skip-gram predicts surrounding words given the current word. $w(t)$ represents t -th word. (Figure from Mikolov et al. 2013a).

Word embedding has proven to be efficient in capturing context similarities. As shown in Fig. 8, words with similar meaning tend to gather nearby in the word embedding space. This is one of the most attractive characteristics of word embedding.

Limitation : Out of vocabulary When using a trained model, words that have never been used in the training can come in as input, due to a unique word such as a name or newly-coined words. These words are usually called unknown words

⁴site: <https://blog.kaggle.com/2016/05/18/home-depot-product-search-relevance-winners-interview-1st-place-alex-andreas-nurlan/>

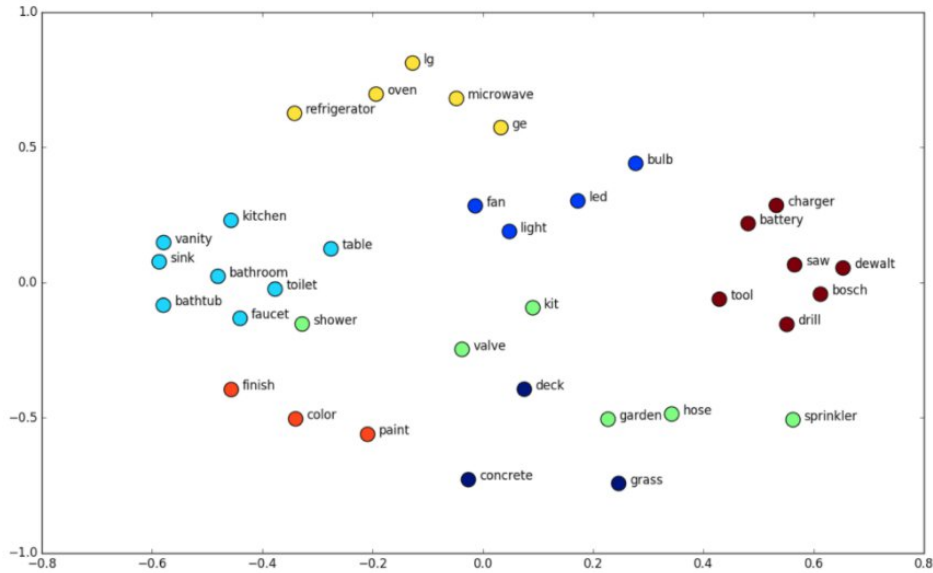


Fig. 8: An illustration of word embedding. Words with similar meaning tend to gather close. (Figure from the site⁴).

(UNK), which are not in the vocabulary at training. One of the limitations of word embedding is that it would not respond well to unknown words, and this limitation is called out of vocabulary (OOV) problem. To cope with this limitation, three methods are largely achieved.

1. simply enlarge size of vocabulary.
2. use copy mechanism for unknown words.
3. use character-level embedding.

The simple way of enlarging the vocabulary can alleviate the limitation, but causes the problem that the amount of calculation becomes too large as the vocabulary size increases. To address this problem that calculation increase, methodologies such as sampled softmax have emerged (Jean et al., 2014).

Second way is use of copy mechanism. When the same words are used in the input and output sentences, we can just copy the same words in the input sentences to output sentences (Fig. 9). That is, when an unknown words come in, the trained model can address the unknown words by simply moving it to the output Gulcehre et al. (2016); Merity et al. (2016). The limitations of this method are expected to be difficult to cope with when the words is only in input sentence and is not in the output sentence.



Fig. 9: An illustration of copy mechanism for machine translation (Figure from Gulcehre et al. 2016).

Character-level embedding Character embedding naturally copes with the unknown word issue. The word is a combination of individual characters. In a language in which the text consists of a combination of characters as well as letters and the meaning of the words corresponds to the synthesis of the characters (e.g. Chinese), character-level system construction is a natural choice to avoid word segmentation (Chen et al., 2015). Therefore, studies applying deep-learning techniques to these languages tend to favor character embedding over words (Zheng et al., 2013). See, for example, Peng et al. (2017) proved that radical based processing can significantly improve sentiment analysis performance.

Word embedding can capture grammatical and semantic information. However, in tasks such as part-of-speech tagging and entity name recognition, the type information within words is also very useful. The natural language understanding of the

character level is of some interest (Dos Santos and Gatti, 2014; Santos and Zadrozny, 2014; Santos and Guimaraes, 2015; Kim et al., 2016).

For certain NLP tasks, improved results for morphologically complex languages are reported. Santos and Guimaraes (2015) applied state-of-the-art levels to Portuguese and Spanish corpus by applying character-level embedding with word embedding to the object name recognition problem. Kim et al. (2016) showed a positive result by constructing a neural language model using only character embedding.

However, as words change into character units, the length of the sentence becomes several times longer. The performance of widely used deep learning models tends to degrade as the sentence becomes longer, thus character-level embedding is not used more than expected. The relationship between out of vocabulary problem and performance degradation as sentence length longer is in a trade-off relationship.

2.2.3 Recurrent Neural Network (RNN)

Recurrent Neural Network (Elman, 1990; Hochreiter and Schmidhuber, 1997; Schmidhuber, 2008; Graves et al., 2012; Sutskever, 2013) is a network that processes sequential information. Unlike traditional neural networks, RNN assumes that all inputs are independent. The term 'recurrent' comes from the fact that the model does the same for each instance of the input sequence, and that the output is dependent on previous operations and results. Generally, a fixed-size vector is created to represent a sequence by entering a single token into the recurrent unit. In this way, the RNN 'memorizes' the previous operation results and utilizes this information in the current operation. These templates can be used in language modeling (Mikolov, 2010; Sutskever et al., 2011), machine translation (Auli et al., 2013; Liu et al., 2014; Sutskever et al., 2014), speech recognition (Graves et al., 2013b; Graves and Jaitly, 2014; Sak et al., 2014), and image captioning (Vinyals et al., 2015b; Karpathy and

Fei-Fei, 2015; Xu et al., 2015). This has led to widespread adoption of RNN in NLP applications over the last few years.

The need for RNN Because the RNN processes data sequentially, it has the ability to capture the sequential nature inherent in the language. The word has a meaning based on the previous word. A simple example of this would be the difference in meaning between 'dog' and 'hot dog'. RNN is designed to model this context dependency and researchers have been a strong motivation to use RNN over CNN. Another factor suitable for this RNN sequence modeling is the ability to model various text lengths, including very long sentences, paragraphs, and even documents (Tang et al., 2015). Unlike CNN, RNN has a flexible calculation step that can capture unlimited contexts. This ability to process input values of arbitrary length has become one of the selling points of major studies using RNN (Chung et al., 2014).

Many NLP tasks also require semantic modeling for the entire sentence. This is related to creating a gist of sentences within a fixed-dimensional hyperspace. These instances are appropriately captured by the RNN. RNN usage has increased for tasks such as machine translation (Cho et al., 2014) where the entire sentence is summarized as a fixed vector and then mapped to a variable-length target sequence.

The RNN also provides network support for time distributed joint processing. Most of the sequence labeling tasks, such as part-of-speech tagging (Santos and Zadrozny, 2014), are based on these domains. More specific use cases are document classification (Li et al., 2015b; Yang et al., 2016; Liu et al., 2016), question answering (Xiong et al., 2016; Lu et al., 2016), and sentiment analysis (Poria et al., 2017; Zadeh et al., 2017; Tong et al., 2017).

Long Short-Term Memory and Gated Recurrent Unit Learning long-term dependencies with RNN is difficult problem. When you try to learn long-term dependencies, you would encounter two main problems, vanishing gradient problem and exploding gradient problem (Bengio et al., 1994). Long short term memory (LSTM (Hochreiter and Schmidhuber, 1997; Gers et al., 1999)) and Gated Recurrent Unit (GRU (Chung et al., 2014)) address these problem. LSTM, as shown in the following illustration Fig. 10, adds gates, and memory cell to a simple RNN, and GRU adds gates to a simple RNN.

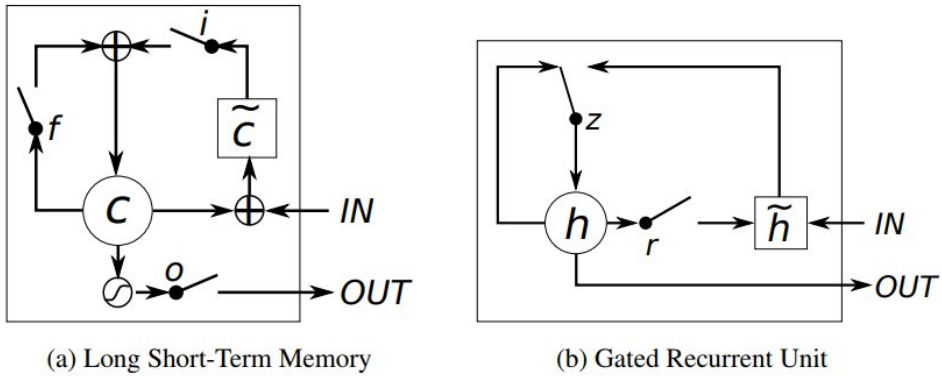


Fig. 10: (a) Long short term memory, i , f and o are the input, forget and output gates, respectively. c and \tilde{c} denote the memory cell and the new memory cell content. (b) Gated Recurrent Unit, r and z are the reset and update gates, and h and \tilde{h} are the activation and the candidate activation. (Figure from Chung et al. 2014)

LSTM consists of three gates, input / forget / output gate, and the memory cell. Hidden state of LSTM is calculated according to the following equation.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2.1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2.2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (2.3)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2.4)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (2.5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (2.6)$$

GRU consists of two gates, reset / update gate. Hidden state of GRU is calculated according to the following equation.

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2.7)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (2.8)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (2.9)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2.10)$$

With this unique mechanism, unlike RNN, LSTM and GRU allow errors to be propagated back to longer time steps. By using LSTM or GRU, you can alleviate vanishing / exploding gradient problem.

2.2.4 Convolutional Neural Network (CNN)

The facts already mentioned are some of the motivations for researchers to choose RNN. However, it is wrong to conclude that RNN is superior to other net-

works. Several recent studies provide evidence that CNN is superior to RNN. CNN achieved a more competitive performance than RNN, even though it is an appropriate task for RNNs such as Language modeling (Dauphin et al., 2016). CNN and RNN have different purposes (functions) when modeling sentences. CNN tries to extract the most important n-grams while RNN tries to generate long boundless sentences. Both are efficient in capturing n-gram features, but the sensitivity to word order is locally limited and long-term dependencies are usually ignored.

Yin et al. (2017) presents an interesting insight into RNN and CNN performance. After evaluating several NLP tasks including sentiment analysis, QA, and part-of-speech tagging, they concluded that there was no "complete winner." The performance of each network depends on the global semantics required by the task. Some of the RNN models used extensively in recent studies are described below.

Deep learning models have recently achieved remarkable results in computer vision (Krizhevsky et al., 2012) and speech recognition (Graves et al., 2013b). Many studies using deep learning, even only in natural language processing, have shown that the word vector representation (Bengio et al., 2003; Mikolov et al., 2013a,b,c; Pennington et al., 2014) vector (Collobert et al., 2011). CNN utilizes layers that wrap around filters that apply to regional features (LeCun et al., 1998). Although the CNN model was originally invented in computer vision, the CNN model has been known to be effective in NLP and can be applied to semantic parsing (Yih et al., 2014), information retrieval (Shen et al., 2014), sentence modeling (Kalchbrenner et al., 2014), and other traditional NLP tasks (Collobert et al., 2011).

Since word embedding has become popular and its performance has been verified, the need for efficient functions to extract high-level features from word combinations or n-grams has increased. These abstracted features can be used for a variety of NLP problems such as sentiment analysis, summarization, machine translation,

and QA. CNN was a natural choice due to its superior performance in computer vision (Krizhevsky et al., 2012; Sharif Razavian et al., 2014; Jia et al., 2014).

CNN has the ability to extract key n-gram features from input sentences to produce a potential semantic representation of the sentence. A pioneering achievement in this area is Collobert et al. (2011), Kalchbrenner et al. (2014), and Kim (2014). These follow-up studies have enabled CNN-based networks to spread widely.

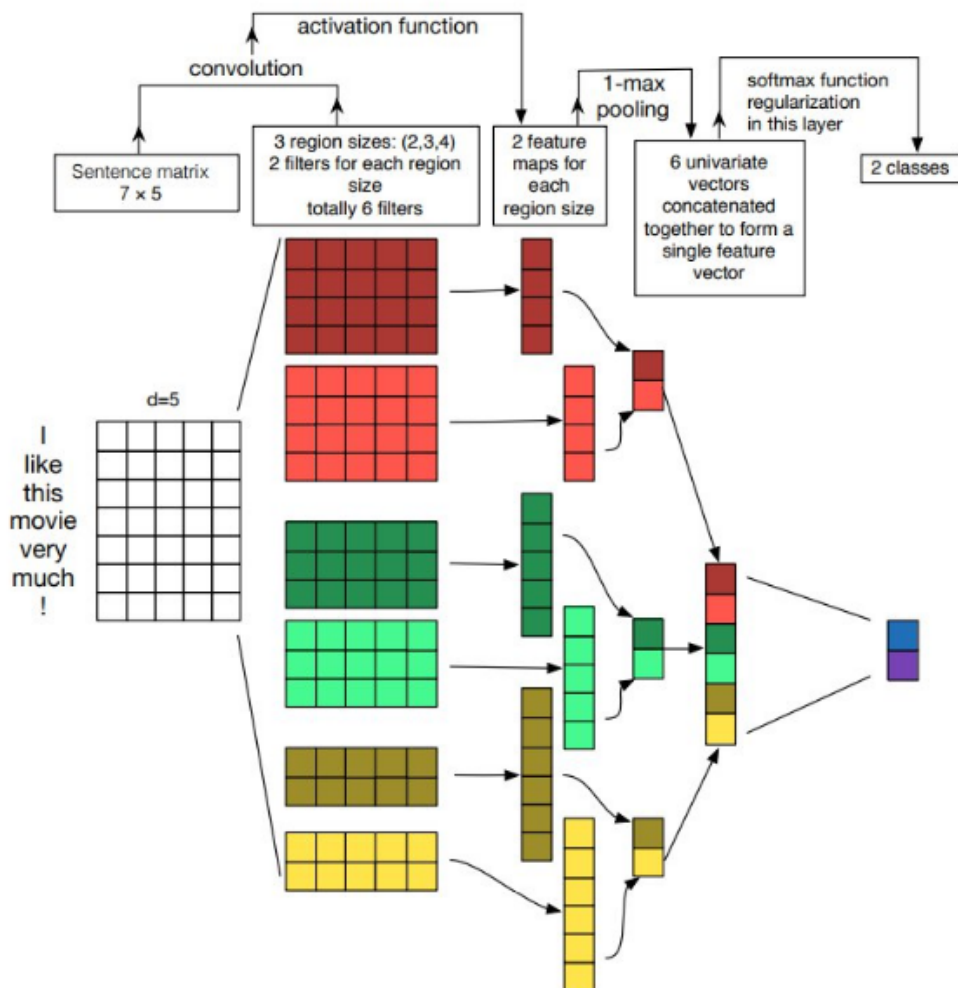


Fig. 11: Convolutional neural network for text processing. (Figure from Zhang and Wallace 2015)

Sentence modeling If an embedding vector corresponding to the i -th word of the sentence is x_i , and the number of dimensions of the embedding vector is d , given a sentence consisting of n words, the sentence can be represented by an $n \times d$ embedding matrix. The following shows the input sentences of the CNN framework.

Let $x_{i:i+j}$ be the concatenation of $x_i, x_{i+1}, \dots, x_{i+j}$. Convolution is performed on this value. Convolution filter w is a vector with dimension hd . This filter is applied to h word vectors. For example, the feature c_i newly extracted by the convolution filter is generated using $x_{i:i+h-1}$. The expression is:

$$c_i = f(w \cdot x_{i:i+h-1} + b). \quad (2.11)$$

Where scalar b is the bias term and f is the nonlinear active function, such as hyperbolic tangent. The filter w uses the same weights to generate feature maps and applies to all possible windows:

$$c = [c_1, c_2, \dots, c_{n-h+1}] \quad (2.12)$$

The number of convolution filters (also called kernels) on CNN is typically hundreds. The widths of the filters are different, and each filter extracts a specific pattern of n -grams.

The convolution layer usually follows the max pooling layer, ie $\hat{c} = \max(c)$. Max pooling sub-samples the input by taking the maximum value for c . There are two reasons for using this strategy.

First, max pooling generally provides a fixed-length output required for classification. Therefore, even when the size of the filter is different, max pooling always maps the input value to a fixed-dimension output.

Second, max pooling reduces the dimension of output while maintaining the

most essential n-gram features in the entire sentence. This is done in an invariant manner because the individual filter can extract a particular feature (for example, 'no') at any point in the sentence and can append it to the final sentence representation.

Word embedding can be randomly initialized or pre-learned in a large, unlabeled corpus. The latter is sometimes useful for improving performance, especially when the amount of correct data is small (Kim, 2014). This combination of convolution layer and Max pooling often overlaps to create a deeper CNN network. This sequential convolution allows to improve the analysis of my sentences by grabbing highly abstracted representations containing rich semantic information. A deep convolution filter (kernel) covers a large part of the sentence until you create a complete summary of the sentence feature.

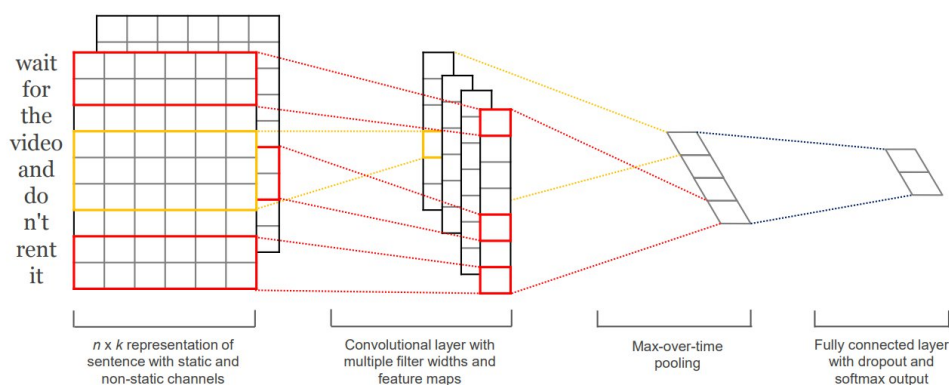


Fig. 12: Model architecture with two channels for an example sentence. (Figure from Kim 2014)

2.2.5 Siamese Network

Siamese Network (Bromley et al., 1994; Mueller and Thyagarajan, 2016; Yin et al., 2015) is a frequently used model for tasks that deal with sentence pairs. Divide

the sentence pair into two sentences, and use each sentence as input to the same model sharing the parameters. The figure below shows Siamese LSTM and Siamese CNN.

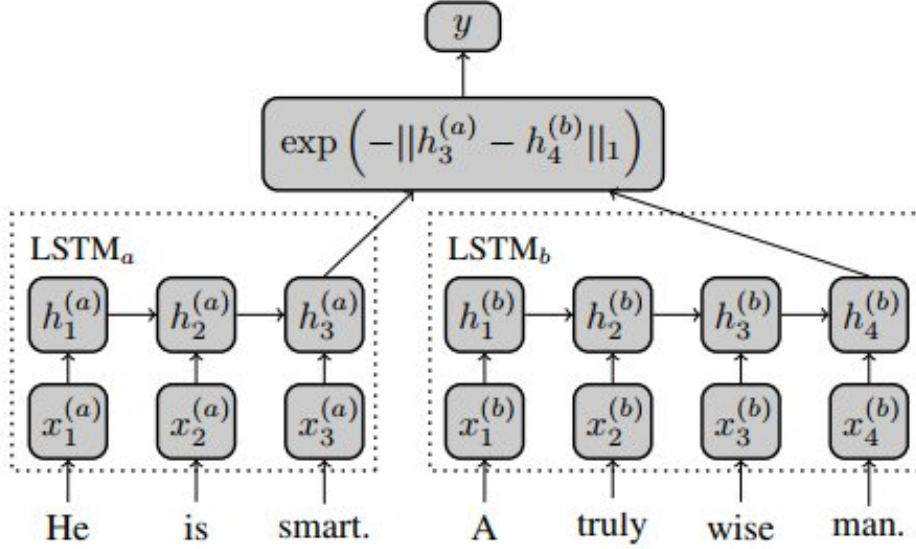


Fig. 13: Siamese LSTM. (Figure from Mueller and Thyagarajan 2016)

2.2.6 Sequence to Sequence (Seq2Seq)

There is the famous model to generate sequence well, called encoder-decoder model or seq2seq (Cho et al., 2014; Sutskever et al., 2014). Two RNN is usually used, one as encoder and the other as decoder (Fig. 15). Encoder make input sentence into a vector, and decoder take the vector as a initial state of RNN. The vector usually is called sentence embedding which represent the sentence information.

2.2.7 Attention Model

To generate sentence better, the concept of attention is introduced (Bahdanau et al., 2014; Vinyals et al., 2015a; Luong et al., 2015; Xu et al., 2015) (Fig. 16).

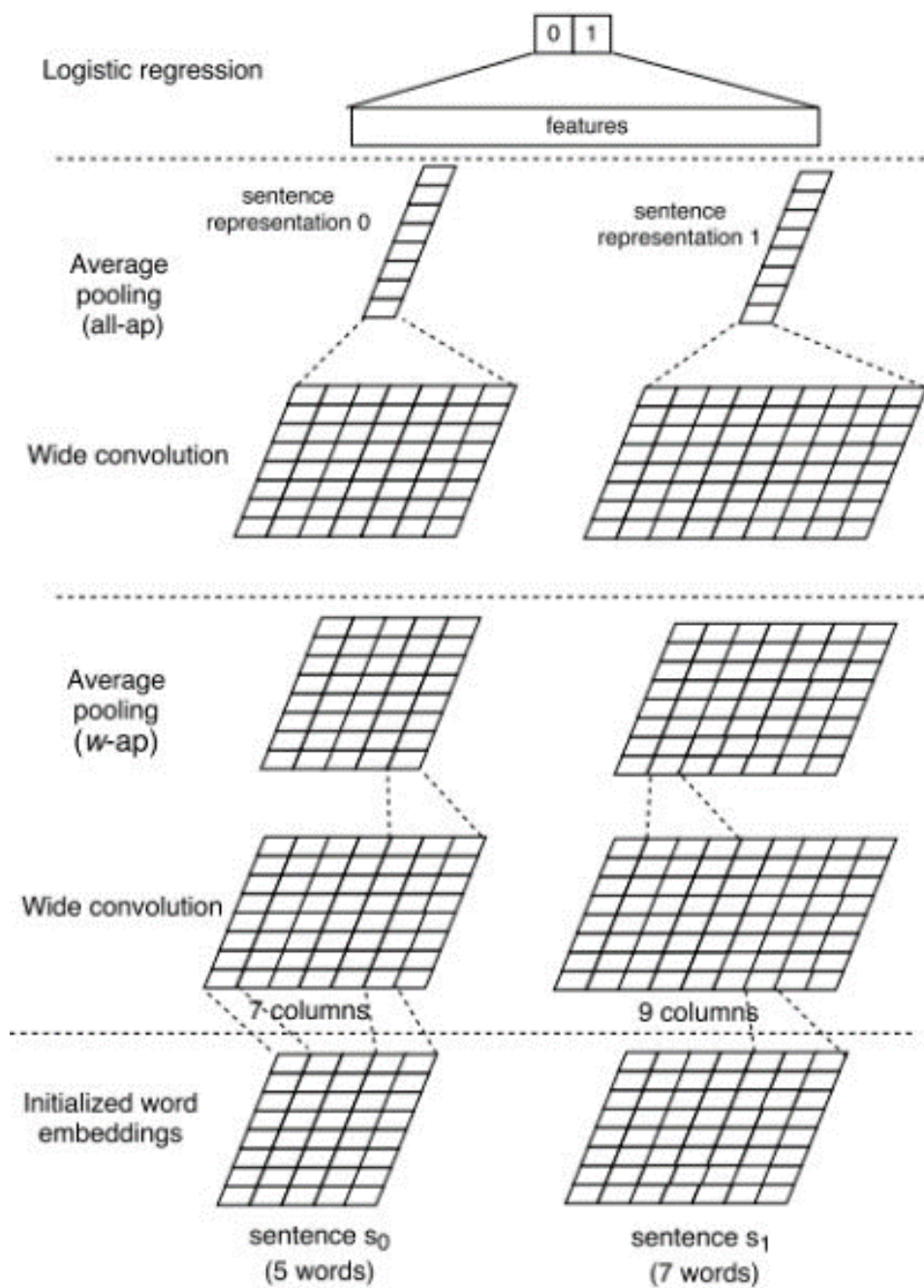


Fig. 14: Siamese CNN. (Figure from Yin et al. 2015)

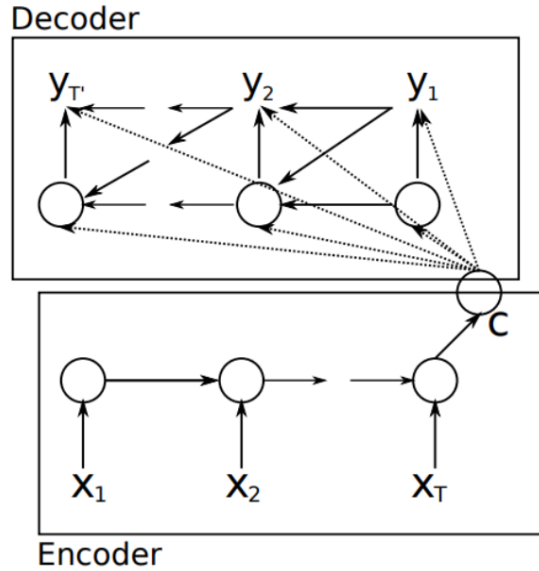


Fig. 15: An illustration of encoder-decoder model (Figure from Cho et al. 2014)

Attention means that the align between input sentence and output sentence. The result of attention, attention matrix as shown in Fig. 17, shows a match of each words in the input and output sentence. This is very useful information to investigate the generation performance of trained model.

co-attention model Attention also can be applied to sentence pair tasks. The attention mechanism for sentence pair is usually called as co-attention (Xiong et al., 2016; Lu et al., 2016). It takes two sentences as input and calculate attention between them which represents the relationship between them (Fig. 18. After that, with the attention information, try to address sentence pair tasks such as answer selection, paraphrase identification, and textual entailment (Yin et al., 2015; Cui et al., 2016). It has some variants, called attention based CNN (ABCNN, Yin et al. 2015), attention over attention (Cui et al., 2016) and so on.

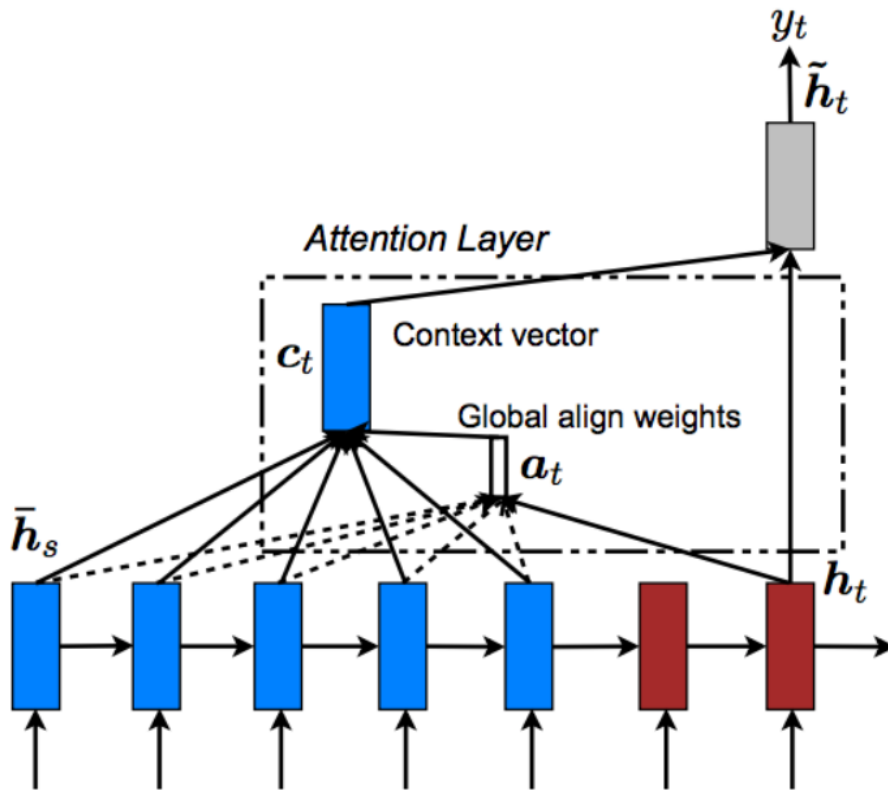


Fig. 16: An illustration of attention mechanism. (Figure from Luong et al. 2015)

self-attentive model Nowadays, attention is calculated between sentence even it-self, called self-attention (Fig. 19). Strictly speaking, it's not the same state of a sentence, but a slightly different state of the sentence. In machine translation task, this method is currently state of the art. It suggests that the attention mechanism has much potential to solve some problems well.

⁵Google research blog: <https://research.googleblog.com/2017/08/transformer-novel-neural-network.html>

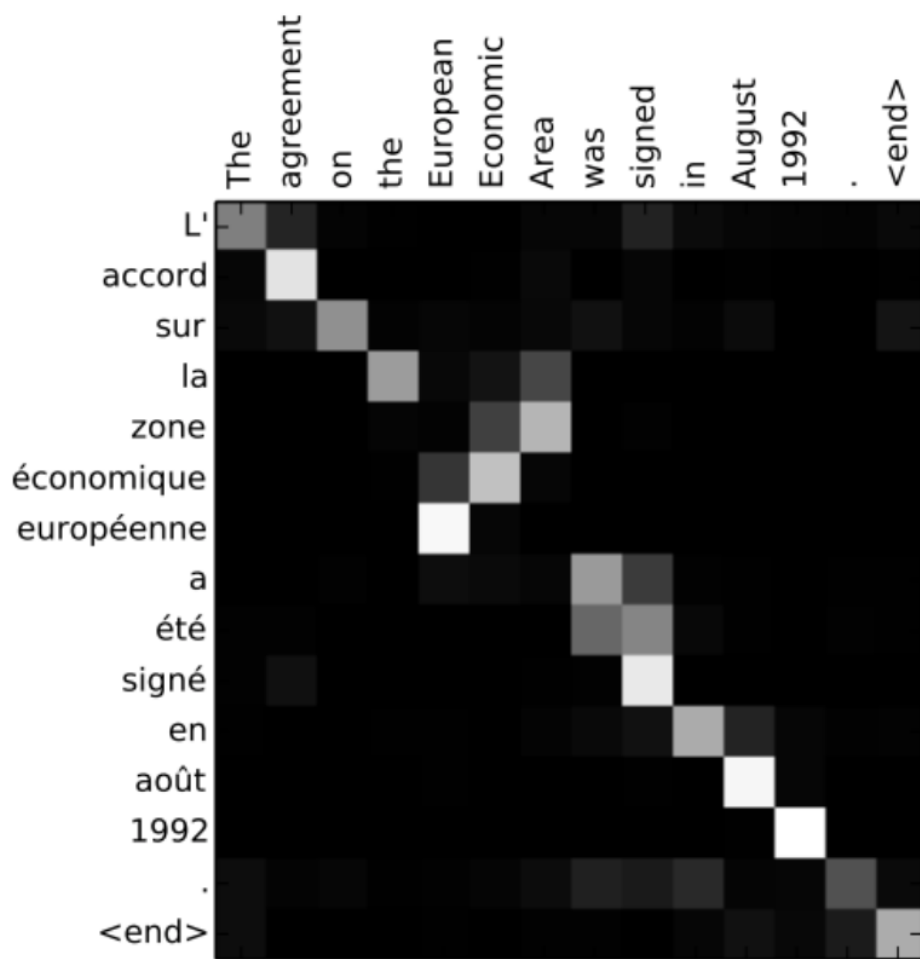


Fig. 17: Example of attention matrix. (Figure from Bahdanau et al. 2014)

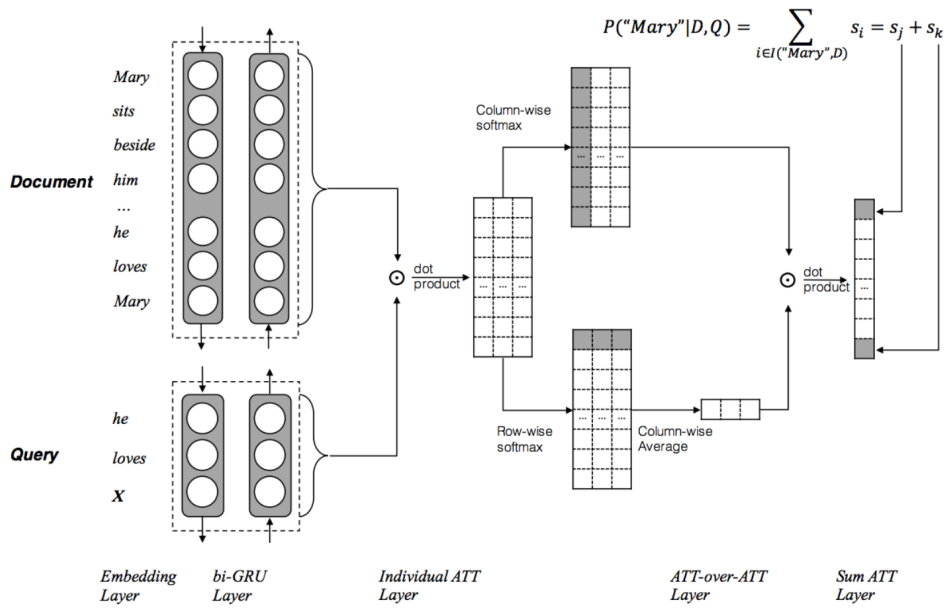


Fig. 18: An illustration of attention over attention. (Figure from Cui et al. 2016)

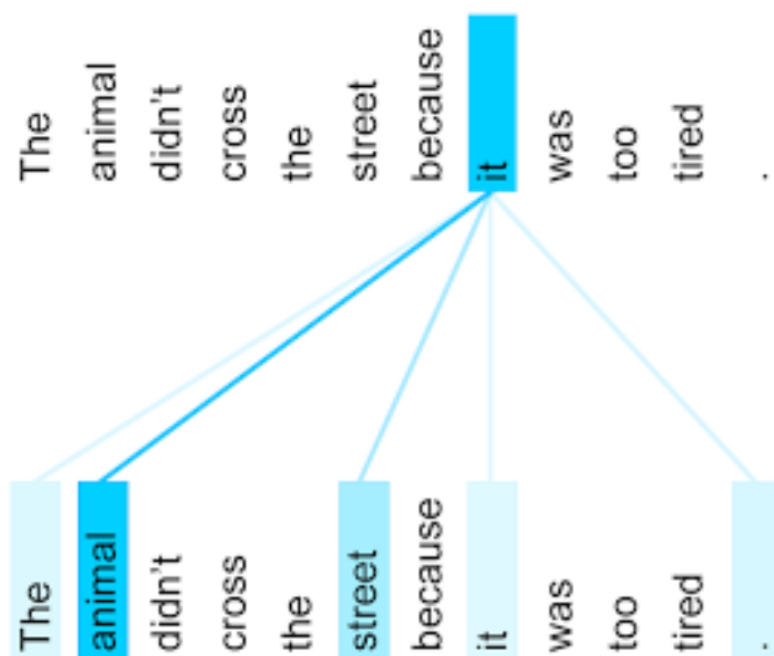


Fig. 19: An illustration of self-attention (Figure from the site⁵)

2.3 Chapter Summary

In this chapter we compared and analyzed the concept of linguistic alignment, and defined alignment / non-alignment criteria. We briefly took a look at deep learning and NLP technology and its model for alignment classification and generation.

Since there was no adequate study of the linguistic alignment criteria, we set standards first. Since preceding alignment studies have pointed to different phenomena, it is meaningful to set standards for them and to classify the conversation. Linguistic alignment refers to the structure of continuing conversations within the context of the previous utterance: Affirms [affirmation] or denies [negation] the words of the preceding interlocutor, and repeats the previous statements [repetition]. Non-alignment refers to cases in which the topic changes suddenly [switch], and conversational flow is interrupted [stop] or completely irrelevant [irrelevant speech].

Linguistic alignment is a phenomenon which can be recognized as analogous, but it is difficult to establish rules for each situation. Therefore, it is not easy to analyze the alignment structure by the rule-based method. So, we tried to utilize the state-of-the-art deep learning method to investigate the linguistic characteristics.

Chapter 3

Experiment

3.1 Overview

This chapter describes data acquisition, alignment classification experiments, aligned response generation experiments and marker validation.

In section 3.2, we made 2 types of the corpus; textbook and twitter corpus by labeling 1,384 sentence pairs of textbooks and 2,000 sentence pairs of tweets. Labeling was performed based on the linguistic alignment criteria in section 2.1. When human determined alignment, the whole sentence was usually taken into account, not just some of the words involved (we will discuss it in more detail in section 3.2.3). To train alignment classification model, a researcher created several sets of inputs and outputs. The model learns how to classify the sentences based on what the researcher tagged as alignment / non-alignment. Through training, a machine makes the similar judgment as what human did. Then, we conducted data augmentation because we had too small and unbalanced data to train deep neural networks models, alignment classifiers.

In section 3.3, we tried to classify sentence pairs into alignment and non-alignment with deep learning classifier and marker-based approach (see Fig. 20). The classifiers were trained with 3,920 pairs in textbook corpus and 4,500 pairs in twitter corpus, after data augmentation. The performance of the classifiers was compared to the baseline marker experiments. The classifiers shows quite good performance (70 – 80%), better than the baseline (35 – 55%).

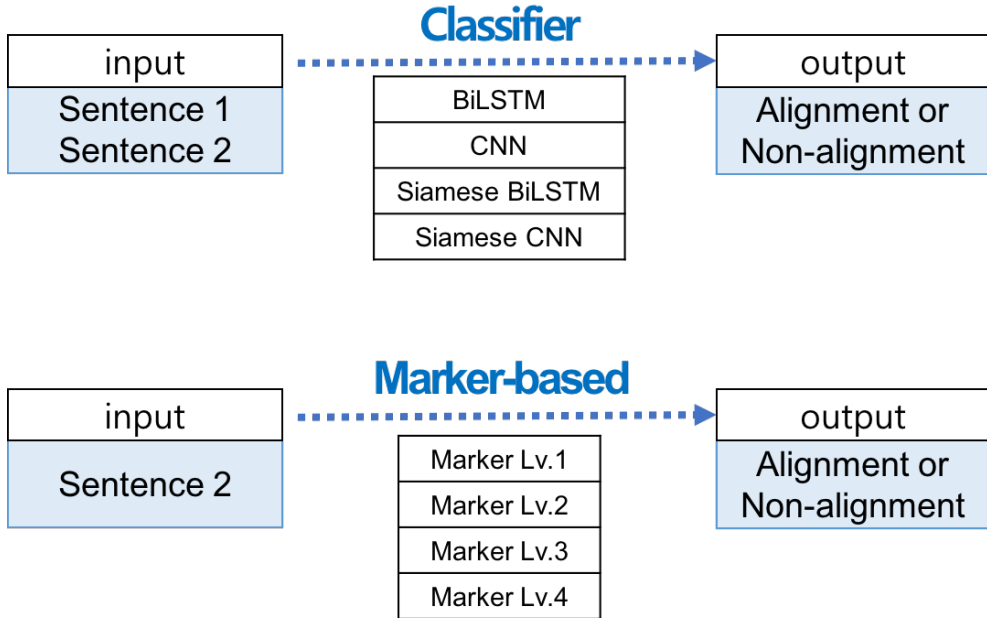


Fig. 20: Alignment classification with deep learning classifier and marker-based approach

In section 3.4 and 3.5, we tried to generate aligned responses with neural machine translation (NMT) model (see Fig. 21).

In section 3.4, we tried to train aligned response generator with 1,226 aligned sentence pairs in twitter corpus, labeled by us in section 3.2). Training, however, did not work out as we expected. We assumed that the training requires more data to produce appropriately aligned responses.

In section 3.5, to solve the lack of data, 50,000 pairs of unlabeled tweets were classified through a trained alignment classifier, CNN. As a result, 27,449 sentence pairs were classified as alignment pairs and we used this as larger alignment corpus. By using the larger alignment corpus, we would train better aligned response generation model and generate more appropriate aligned response to the user. It was possible to get a better performance and quite good results from the training with the

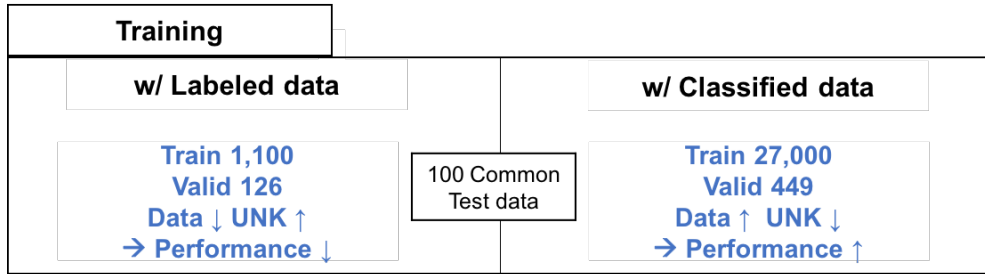
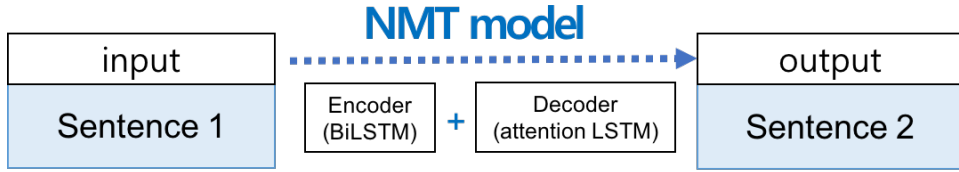


Fig. 21: Aligned response generation with NMT model

classified 27,449 aligned sentence pairs than with the labeled 1,226 aligned sentence pairs. This means that by using the trained classifier, the data can be acquired with ease and can reduce the human labor cost.

The more labeled data, the better results. To get numerous data conveniently, we used a high-performance classifier which could replace human labor. The models classify unlabeled new sentences based on the correctly labeled answers made by human researchers. Using classified sentences reduces the cost of human effort to label sentences, and the whole procedure becomes less costly.

In section 3.6, we validated Spanish markers based on the linguistic human criteria and were able to find interesting and meaningful markers in the classified 27,449 aligned sentence pairs, different from those expected by humans. In other words, the computer learned human-like ability to judge aligned sentence pairs. We used the deep learning method to imitate human judgment without any rules. This suggests

that obtaining a large corpus by the trained classifier and then analyzing the large corpus classified by the trained classifier could be a useful language analysis method.

3.2 Data Acquisition

This study is based on the existing natural language processing techniques, and it implements and validates the deep learning model based on the latest research.

3.2.1 Data Types

We mainly analyze conversational data which are composed of dialogue, which are written in colloquial tone, called quasi-spoken corpus. The source of data that can be obtained directly include 4 collections of Spanish language learning materials at Seoul National University, 3 scripts for foreign language theatrical performances, and Spanish tweets crawled from web. Due to privacy protection and accessibility limitation, other SNS or posts were limited to use, while Twitter was relatively easy and available.

At the beginning of the study, play script data was also used, but in the case of scenario was removed due to copyright problems. All dramas are very famous Spanish works, and were used as a script for the foreign drama festival at Seoul National University: *Prohibido suicidarse en primavera*, *Siete gritos en el mar*, and *Morir (o no)*. The play *Prohibido suicidarse en primavera* was premiered in Mexico in 1937 and later was also produced as a TV soap opera. *Siete gritos en el mar* was performed in Argentina from 1952 to 1968 and was made into a film in 1954. *Morir (o no)* is a play written by Sergi Belbel, and was also filmed in 2000. He won the award for this play in 1994 and 1996, respectively.

In addition to the refined data from textbooks, it also has the advantage of being able to obtain more vivid dialogue that is often used on the web.

Textbook A total of 4 textbooks are used. They were published in 2013 for undergraduate students at Seoul National University: Beginner and Basic, Intermediate,

Table 12: Data composition before adjustment

	Alignment	Non-alignment	Total	Ratio	Average length	Vocab size
Textbook	980	264	1,244	3.7:1	20.85 (10.43)	2,887
Twitter	1,226	574	1,800	2.1:1	28.67 (14.34)	6,322

Conversation and Writing (published in 2008) textbooks (Shin et al., 2013a,b; Macías et al., 2013; Shin et al., 2008). Each book is composed of highly refined dialogues. It is useful for learning situational and specific phrases because the chapters are topic-specific. Most of the conversations are an alignment structure.

Twitter Textbook data are suitable for learning the alignment structure for specific topic, but were not suitable for obtaining a sufficient amount of non-alignment data. In order to make a model to train the patterns on a computer, Twitter data was suitable in a highly flexible syntax. Twitter is known as a good linguistic resource that can be freely and automatically collected without restrictions.

3.2.2 Crawling and Preprocessing

(1) Firstly, need to be approved on twitter for crawling¹.

- Open <https://apps.twitter.com/> and click the button: ‘Create New App’
- Once the app is created, you will be redirected to the app page.
- Open the ‘Keys and Access Tokens’ tab.
- Copy ‘Consumer Key’, ‘Consumer Secret’, ‘Access token’ and ‘Access Token Secret’.

¹Reference site: <http://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>

- (2) Crawl the tweets with more than 100 times RT(retweeted) on twitter and its comments.
- (3) Preprocessing : Erase unnecessary information from the crawled data and use a tokenizer.

Due to the nature of Twitter, misleading or unscripted expressions and unrefined sentences are mixed. To crawl proper conversations among these sentences, we needed to select Spanish sentences that were mentioned many times by the users. Thus, we picked out the tweets with more than 100 times RT and its comments.

We removed account(accountname), hashtag (#something), 'RT', links, special characters to clean tweet text. The python library 'Regular Expression' is utilized for preprocessing.

We used ToktokTokenizer, a tokenizer for Spanish. When using ToktokTokenizer only, the 'word' + '.' sometimes are not be separated properly. For example, 'done.' should be separated into 'done' + '.' but they were not separated. Thus we additionally used `nlk.wordpunct_tokenize`.².

3.2.3 Data Labeling

Although we had the colloquial data for research such as Spanish textbooks, play scripts, and Twitter data, it is difficult to acquire the labeled data to train. Therefore, we should attach labels to sentence pairs by ourselves. Assuming conversational communication situations, labels were attached to the sentence pairs in which neighboring responses exist.

Alignment The data input and output values used in this study are as follows.

²Reference site: <https://github.com/nltk/nltk/issues/1558>

- Input: sentence 1, sentence 2 — Output: Alignment / None

We did not judge based on specific words for labeling. Instead, we distinguished by the meaning of the whole sentence. The reason for this is that judging based on only a specific word can artificially affect the experimental results, and the meaning of the actual conversation can not be represented by only a few words. In order to distinguish between alignment and non-alignment based on the meanings in the whole sentence, the entire sentence was entered to avoid overlooking or omitting any elements.

Sentence 1
= previous sentence

Sentence 2
= following Sentence
= aligned response

Diálogo

En el estadio de su universidad: Pedro y Carla hablan de sus gustos

Carla:

Hola, Pedro. ¿Cómo va todo?

Pedro:

¡Feliz! Acabamos de ganar el partido de fútbol.

Carla:

¡Enhorabuena! ¿Juegas todos los días al fútbol?

Pedro:

Sí, me encanta jugar al fútbol. ¿Y, a ti, qué deporte te gusta practicar?

Carla:

A mí me gusta jugar al baloncesto, pero también me gusta ver el fútbol. Me emociona ver los partidos de la Copa Mundial.

Pedro:

Sí, son emocionantes.

Carla:

Aparte del fútbol, ¿qué otras cosas te gustan?

Pedro:

Bueno, me fascina escribir poemas y también tocar el piano.

Fig. 22: Selection of alignment in textbook (*Español Básico 1*, Chpt 10, p.172)

3.2.4 Data Augmentation

Deep learning models require more than thousands of data. We have relatively small data. So we try to supplement the amount of data by doing a data augmentation.

According to a forum³, currently there is no effective text data augmentation technique. Among very few candidates, data augmentation using thesaurus is most

³Forum site: <http://forums.fast.ai/t/data-augmentation-for-nlp/229>

Sentence 1
= previous
sentence



Fig. 23: Selection of alignment in Twitter

effective. However, there is no thesaurus available for Spanish, and since it was costly to make it, we look for another way.

Another simple augmentation technique is to increase the number of sentences by making each word of the sentence into an unknown token (UNK) with a certain probability. This method is called word dropout (Dai and Le, 2015; Bowman et al., 2015; Xie et al., 2017). There are two main ways; one is to change a certain percent of words in a sentence into UNK, and the other is to change each word into UNK with a certain probability. Since we could make more patterns by changing each word (increased by $2^{|Number of Words|}$), we used the word-by-word method in this study. This method is known to increase the generalization capability of the model by preventing the model from overfitting to small train data.

Table 13: Examples of data augmentation using word dropout

Alignment	Non-alignment
<ul style="list-style-type: none"> ¿Cuándo es tu <unk>? / Es el quince de <unk> 1 ¡Son <unk>! / ¡Así es! 1 Perdón por <unk> con mis problemas. / Ni hablar. No te preocupes. 1 Carlos, ¡gracias por <unk> / ¡De nada! Para eso están los amigos. 1 ¿De dónde <unk>? / De España. 1 	<ul style="list-style-type: none"> Bueno, <unk> está la gente del pueblo? Hace tiempo <unk> no sé nada de ellos. ¿Te acuerdas de <unk> la chica cuyo <unk> tenía <unk> zapatería cerca de la plaza? 2 Estaría llamándote a ti. Pero ¿por qué no llamaste después? / Tienes <unk> ¿no? 2 Porque <unk> aplicación <unk> de mensajería por Internet ha aumentado en todo el mundo. / ¿Y quién sale perjudicado? 2

Cross-validation and test data in deep learning In the traditional machine learning field, cross-validation was usually used, especially when the data size is hundreds (Kohavi et al., 1995) (Fig. 24). Cross-validation, however, is not often used in deep learning⁴. Cross-validation usually takes more time, at least five times longer to perform one experiment, and each deep learning experiment usually takes over an hour. Thus, using cross-validation is time-consuming and inefficient in deep learning field.

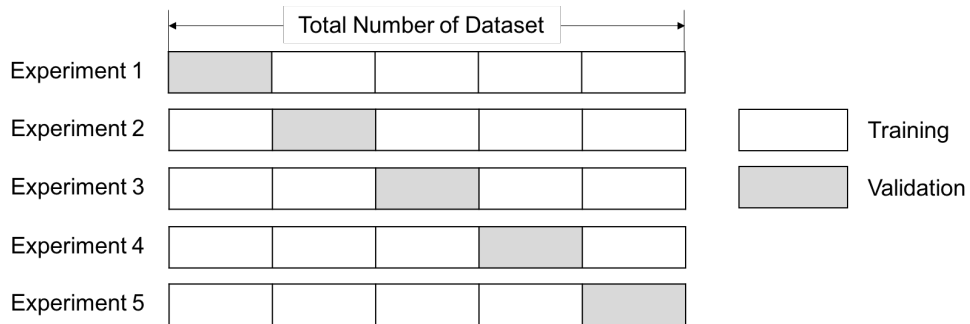


Fig. 24: An illustration of cross-validation

Instead of cross-validation, many researchers tend to train the model with valid

⁴Quora Q&A: <https://www.quora.com/Is-cross-validation-heavily-used-in-deep-learning-or-is-it-too-expensive-to-be-used>

loss for model selection (called early stopping Prechelt 2012, Fig. 25). The train data divided into a training set is used to train the model and a validation set is used to estimate model performance during training. A validation set is never used for model training, and valid loss is calculated from a validation set. Deep learning researchers usually make previously a test set that reflects the target performance. We also made a selected test set (See chapter 3.3). In this study, 10% of train data was used as validation set. With the premade test set, model performance was assessed.

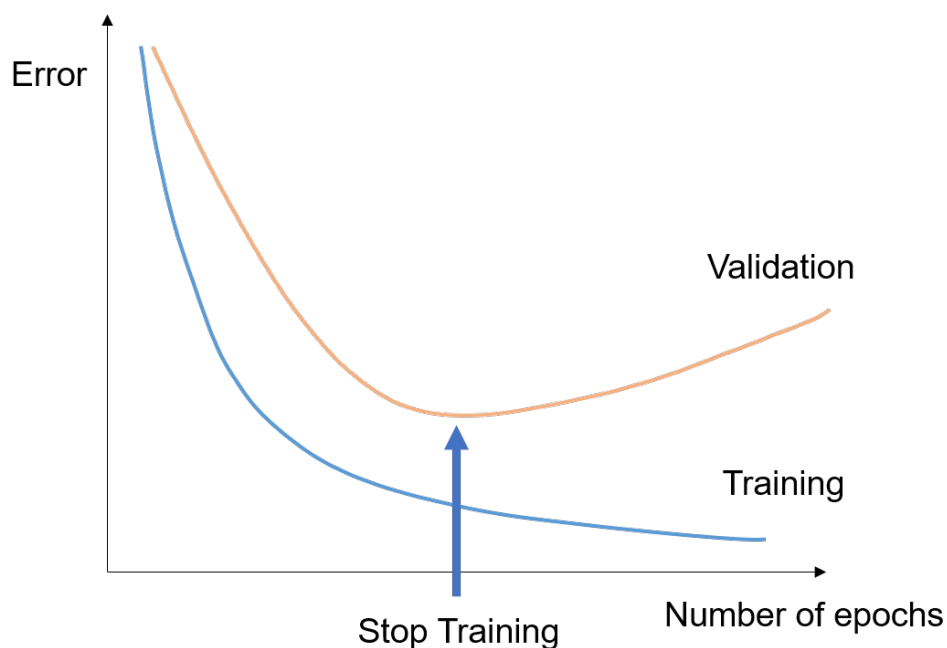


Fig. 25: An illustration of early stopping

3.3 Alignment Classification

Before the generation of aligned responses, we assumed that we should identify alignment sentence pairs to collect them for training. Since it is difficult for a person to identify every sentence, this study designs an automatic classifier to reduce high cost for identification of alignment sentence pairs. Based on the trained classifier, we could identify and collect alignment sentence pairs.

Hypothesis 1 Spanish has a systematic lexical structure, and there will be many markers for alignment sentences compared to those that are not.

Hypothesis 2 Markers will play a significant role in automatically determining whether or not to align.

This study introduces deep learning method to dialogue analysis, linguistic alignment, based on existing NLP theory. Recently, various model designs and techniques are blooming in natural language processing field. In this paper, we utilized the meaningful deep learning models and techniques applied to a large number of NLP tasks, and compared the models to each other by applying them to the linguistic alignment.

3.3.1 Baseline - marker-based classification

Marker-based (rule-based) classification was used as baseline approach. We created a marker list first. If any of the words in the marker list appear in the second sentence (following sentence), the second sentence is classified as the aligned response. We called this approach as marker-based (rule-based) classification.

There are very few studies on Spanish markers, especially in the dialogue model. Here we proceed with an empirical validation of the alignment phenomenon on the

Table 14: List of 20 components for verification experiments consisting of 4 levels

List of 20 elements			grammatical function
Lv.1	el	the	article (sg.masculine)
Frequently used anaphors	la	the, it	article (sg.female)
	los	the, them	pronoun, article (pl.masculine)
	las	the, them	pronoun, article (pl.female)
	lo	the, it	pronoun, article (sg.masculine)
Lv.2	sí	yes	affirmation
Positive/negative expressions, frequently used adverbs	bien	well	adverb
	si	if; *yes (online)	conjunction
	mucho	much, many	adjective, adverb
	no	no, not	negation
Lv.3	con	with	preposition
Prepositions, presence statements / obligatory expressions	a	to	preposition
	en	in	preposition
	por	for	preposition
	hay	there is; should	auxiliary verb
Lv.4	siempre	always	adverb
Conjunctions and adverbs that add or emphasize contents	nunca	never	adverb
	y	and	conjunction
	pero	but	conjunction
	así	thus, thereby	adverb

Twitter social media conversation data.

The markers were classified into four stages in the order of significant features and semantic factors. We use 7 types of markers from Danescu-Niculescu-Mizil (2012), which are considered to be processed by humans in a generally non-conscious fashion: articles (*el, la, los, las*), auxiliary verbs (*hay*), conjunctions (*y, pero, si*), high-frequency adverbs (*bien, siempre, nunca, así, mucho*), pronouns (*lo, la, los, las*), prepositions (*con, a, por, en*), and quantifiers (*mucho*).

In the first step, definite articles *el, la, los, las* and object pronouns *lo, la, los, las* are selected as the representative terms. In the second level, we select affirmative/negative expressions to the previous speech and frequently used adverbs. Step

Table 15: Results of baseline experiment

	Types	Accuracy (%)	
		Textbook	Twitter
Lv.1	Anaphors (pronouns, definite articles)	32.14	50.00
Lv.2	Positive/negative expression, Adverbs	40.00	52.50
Lv.3	Prepositions, Obligatory verbs	34.29	50.00
Lv.4	Conjunctions, Adverbs	37.14	53.00

3 consists of some principal prepositions containing detailed explanations of previous dialogues and auxiliary verb hay, which represents both the meaning of presence and duty. Step 4 contains elements that enrich the previous conversations, providing additional contents or emphasizing something.

Again, If any of the words in these step of marker list appear in the second sentence (following sentence), the second sentence is classified as aligned response. We tested baseline approach with selected test data (described in section 3.3.2). As the step progresses, markers of each step are cumulated. In other words, the accuracy is evaluated about whether any of each word belongs to the step: 5 elements in step 1, 10 in step 2, 15 in step 3, and 20 in all. The results are shown in Table 15. It is contrary to the hypothesis 2 that we assumed before the study. That is, the results are about 50-60%, which means the vocabulary elements predicted to be important are not functioning as alignment markers.

3.3.2 Settings

There are four types of deep learning models used in this study. Above all, the model is simplified and abstracted at Fig. 26. Each model (BiLSTM, CNN, Siamese BiLSTM, Siamese CNN) was implemented with Windows 10, python 3.5.4, pytorch 0.3.0, GPU-GeForce GTX 1070.

Selection of test data To investigate model performance properly, we previously selected appropriate test data. In Section 2.1, we divided six subcategories in detail. Linguistic alignment classification can be a much more difficult problem than sentiment analysis, which is a representative binary classification problem among NLP tasks. The alignment statement includes not only a positive response, but also a negative response, such as the criteria previously mentioned. Repeating or re-mentioning the previous topics is also included in the alignment.

Table 16: Alignment for directly related interpersonal conversations

Dialogue		
A: Voy a llamarle por teléfono. (En. I'm going to call him on the phone.)	B: Ya lo he intentado hace unos minutos. Por eso, si le llamas, no te contestaría. (En. I already tried a few minutes ago. So, if you call her, she'll not answer.)	Alignment
B: ¡Mira! ¡Ahí viene ya, viene corriendo! (En. Look! There she comes, she's running!)	C: Siento llegar retrasada ... (En. I'm sorry I'm late ...)	Non-alignment (Reason for exclusion: Not appropriate as an answer from a listener participating in a direct conversation)
C: Siento llegar retrasada ... (En. I'm sorry I'm late ...)	A: ¿Qué te pasó? ¿Cómo es que llegas tan tarde? (En. What happened? How come you're so late?)	Alignment

And, in addition to the non-alignment criteria mentioned in Chapter 2.1, we exclude some sentences, since we are targeting a conversation between two people who are directly related. For example, in the dialogue of Table 16, second row, the sentence "Siento llegar retrasada ... *I'm sorry I'm late*" is not the wrong sentence. But it is not the aligned response for a interlocutor who has heard "¡Mira! ¡Ahí viene ya, viene corriendo! *Look! There she comes, she's running!*" Therefore, it is judged to be a non-alignment. A more appropriate response to B's utterance would be "¡De verdad! ¡Me alegra que haya venido! *Yes! I'm glad that she comes!*" or so. There are many complicated situations like above example.

There is another example; even if it is the same responses, whether it is aligned response or not depends on the situation. The simple example is 'hello'. If you re-

spond 'hello' to the greeting, that is aligned response. If you reply 'hello' to someone who is asking about 'where are you living?', that is non-aligned response.

Considering such points, we made a test set by selecting sentences that need to be classified well. The selected test data is in the following Tables from 17 to 28. 140 pairs (70 alignment pairs and 70 non-alignment pairs) for textbook corpus and 200 pairs (100 alignment pairs and 100 non-alignment pairs) for twitter corpus are selected as test data.

Table 17: Examples of alignment-labeled data: Textbook (1)

Sentence 1	Sentence 2
¿Cómo te llamas?	Me llamo María Pérez.
¿Y tú? ¿Cuál es tu nombre?	Mi nombre es Pedro García.
Encantada de conocerte.	Igualmente.
María, ¿cuál es el correo electrónico del profesor López?	A ver. ... ¡Ah, sí! Su correo es lopez@gmail.com.
A ver. ... ¡Ah, sí! Su correo es lopez@gmail.com .	¿Cómo se escribe?
¿Cómo se escribe?	Se escribe L-O-P-E-Z.
¡Muchas gracias!	¡De nada!
¡Hasta otro día!	¡Vale! ¡Hasta pronto!
¡Vale! ¡Hasta pronto!	¡Adiós!
Buenos días, profesor Moreno. ¿Cómo está usted?	Muy bien, gracias. ¿Y usted?
Muy bien, gracias. ¿Y usted?	Bien, gracias. ¡Hasta luego!
Bien, gracias. ¡Hasta luego!	¡Adiós!
Buenas tardes, señora.	Buenas tardes, doña María. ¿Cómo va todo?
Buenas tardes, doña María. ¿Cómo va todo?	No muy bien.
No muy bien.	¿Qué pasa?
¿Qué pasa?	Estoy cansada.
¡Hola, Paco! Buenas noches. ¿Qué tal?	Bien.
¿Y tú?	Así, así.
¡Chao!	¡Chao! ¡Hasta pronto!
Muchas gracias, señor.	De nada, señorita.
¡Adiós!	¡Hasta la vista!
¿Cómo te llamas?	Me llamo Esther.
Me llamo Esther.	¿Y tu apellido?
¿Y tu apellido?	Kim.
¿Cómo se escribe?	K de kilo, I de iglesia, M de Madrid.
¿Cómo se llama usted?	Mi nombre es Raúl.
Este es mi amigo Juan.	¡Mucho gusto!
¡Mucho gusto!	¡Igualmente!
¡Mire! Le presento a la profesora Claudia.	¡Encantado!
¡Encantado!	¡Encantada!
¡Hola, Sumi!	¡Hola, Carlos!
¿Qué tal?	Muy bien, gracias.
¿Y tú? ¿Cómo estás?	¡Genial!
¿Qué tal tu clase de español?	¡Excelente!
En mi clase hay muchos estudiantes inteligentes.	¡Qué interesante!

Table 18: Examples of alignment-labeled data: Textbook (2)

Sentence 1	Sentence 2
¿Hay aire acondicionado en tu aula?	¡Por supuesto! También hay un ordenador nuevo y un proyector muy moderno, una pizarra verde muy larga y muchas tizas de colores.
¡Por supuesto! También hay un ordenador nuevo y un proyector muy moderno, una pizarra verde muy larga y muchas tizas de colores.	¿De qué colores?
¿De qué colores?	Azules, blancas, rojas y amarillas.
¡Sumi! ¿Qué grande es tu mochila? ¿Qué hay?	Hay un cuaderno verde, un cuaderno de color naranja, lápices, bolígrafos y un ordenador portátil nuevo.
Hay un cuaderno verde, un cuaderno de color naranja, lápices, bolígrafos y un ordenador portátil nuevo.	¡Muchas cosas!
Bueno, hasta pronto, Sumi.	¡Hasta otro día, Carlos!
¿Qué hay en el aula?	En el salón de clase hay una pizarra, unas sillas y unos libros.
¿Qué hay en el salón de clase?	En el aula no hay reloj.
¿Hay una computadora?	Sí, hay una computadora.
¿Hay una computadora?	No, no hay ordenador.
¿Hay un ordenador?	Sí, hay una computadora.
¿De qué color es el libro?	Es rojo.
Aquí la tiene... ¿Qué van a comer?	Para mí, una ensalada de pollo sin cebolla y, de segundo, filete de ternera con patatas fritas.
Para mí, una ensalada de pollo sin cebolla y, de segundo, filete de ternera con patatas fritas.	Yo quiero, de primero, un gazpacho y, de segundo, pollo asado con arroz.
Yo quiero, de primero, un gazpacho y, de segundo, pollo asado con arroz.	¿Y para beber?
¿Y para beber?	Para beber, una botella de agua mineral sin gas.
¿Qué tal? Perdona, ¿cómo te llamas?	Mi nombre es Enrique, pero mis amigos me llaman Quique.
¿Eres español?	No, no soy español. Soy de México.
No, no soy español. Soy de México.	¿De qué parte de México eres?
¿De qué parte de México eres?	Soy de México, D.F. (Distrito Federal). ¿Y tú?
Soy de México, D.F. (Distrito Federal). ¿Y tú?	Yo soy de Sevilla. Soy española.
¿Cuál es su profesión?	Soy ingeniero electrónico.
¿Cuántos idiomas habla?	Hablo tres idiomas: español, francés e inglés.
¿Cuáles son sus cualidades?	Soy trabajador, sincero, entusiasta, amable y puntual.
¿Cuál es su nacionalidad?	Soy chileno.
Muchas gracias por su tiempo, Sr. Pérez.	A usted por la oportunidad.
¿Eres Juan?	No, no soy Juan. Soy Pedro.
¡Bienvenida a Madrid, Mison!	Muchas gracias, señora Pérez.
¿Qué tal el viaje?	Bien, pero estoy un poco cansada.
Es un viaje bastante largo.	¡Sí, es verdad!
Mison, esta es tu habitación. El cuarto de baño está a la derecha de tu habitación y el comedor está al lado de la cocina.	Muy amable, señora.
Una pregunta. ¿Hay una farmacia cerca de la casa?	¡Claro que sí! Hay una farmacia muy grande enfrente de la casa.
Y la universidad, ¿dónde está?	La universidad está un poco lejos, pero hay una parada del bus universitario delante de la casa.
Una última pregunta. ¿Hay un banco por aquí?	Hay muchos bancos dentro del centro comercial del barrio. El centro comercial está entre la oficina de correos y la plaza.
¿Qué hora es?	No sé. Serán las diez y media.

Table 19: Examples of alignment-labeled data: Twitter (1)

Sentence 1	Sentence 2
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IPHONES 8??? ¿Qué esperas? Correleeee! http...	@IamKatieAngel Claro que si, me di cuenta por en video y en cuanto lo vi me puse a hacer todo
RT @flor_vigna: Que dicen 🤔???	@flor_vigna Obviamente todo te queda perfecto
@PabloLanusse: La justicia no es venganza ni morbo. Uno es inocente hasta condena firme. Pero prisión preventiva es válida para garantiz...	@PabloLanusse Tal cual . De acuerdo . Así debe ser .
@AntonioJSMazuec: ¿Cómo habéis empezado la semana familia?	@AntonioJSMazuec La he empezado increíble, gracias a tu me gusta soy la mas feliz del mundo ahora mismo. Gracias...
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal Me gustaría saber cómo hace para usar esta cuenta ud. Desde donde esta.
@_anapastor_: La fiscalía belga respondiendo a todas las preguntas...igual que aquí!!!	@_anapastor_ Esto lo tenemos en cuenta? Es lo mismo con la gente de su país, me gustaría...
@ghanito: España no tiene salvación posible	@ghanito @SenoritaPuri A mi sin embargo me encanta nuestro espíritu burlón y nuestra capacidad de reírnos de casi todo.
@Camila_Cabello: Te amo.....	@Camila_Cabello Te amo más
RT @TNFamosos: Mica Viciconte se animó a un primerísimo primer plano sin maquillaje https://t.co/YJAK2Kpe3c	@TNFamosos Es preciosa y natural se muestra tal y como es por eso me encanta
RT @IgorMeltxor: ALERTA: nazis de caza por Barcelona aplicando su 155.#L6ARVCat #RepúblicaCatalana #IndependenciaCatalunya #NoPassaran #se...	@IgorMeltxor ese se ha llevado un 155 en toda la cara me encanta, si no fuera un indepenfacha no le hubiese pasado nada
@LuisFlorido: Estamos en EEUU, mañana nos reuniremos con Presidente del TSJ en el exilio y magistrados legítimos ¡atentos!	@LuisFlorido a mi me encanta los EEUU . lastima q ya no puedo ir. Este gobierno me secuestró o los venezolanos todos estamos secuestrados
@BeatrizTalegon: La población Catalana tiene una cultura democrática muy por encima de la media española. Por eso aquí no se entiende	@BeatrizTalegon La línea democrática no se...por la intelectual da penica...Por eso aquí pensamos que sois retrasados.
@itsjoelpimentel: Buenos días	@itsjoelpimentel Hola buenos días
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buenos días, Cómo estás?
@luisrodriguezrm: El pelotudo de Beto Casella (que tiene como 70 Pirulos y se hace el pendejo), ataca a Jorge Lanata y defiende al delin...	@luisrodriguezrm Dijiste lo que pienso yo también . Exacto
RT @DANIROVIRA: ¡Qué gran noticia! 🙌🇺🇦❤️ https://t.co/VjgQDfEGKE	@DANIROVIRA Que bien, ojalá que se consiga 🙌🇺🇦
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde Si yo también lo recuerdo siempre que veo sufrir a una persona humilde lo recuerdo, a El, a su madre, hermana, esposa e hijos.
RT @villalobossebas: te amo 🥰	@villalobossebas Yo tambien te amo
RT @marley_ok: Nació Mirko!!! 17:26 horas de acá, 51 cm, 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Bienvenido Mirko! Que tengas una larga vida llena de paz, luz y amor con tu querido papi. Felicitaciones Marley
RT @yuyacst: Te amo.	@yuyacst Te quiero linda pequeña...
RT @ergasto: Yo estoy a favor de que de una vez por todas la Revista Barcelona se independice del Estado Argentino.	@ergasto Yo tambien!!!!
RT @guillercastello: La provincia remueve a 6 funcionarios involucrados en la megaestafa al IOMA que denunciamos junto a @elisacarriohttp...	@guillercastello @elisacarrio Bien hecho
@botedeltitanic1: SI ME AUMENTAN EL ABL NO QUIERO VER MAS EXTRANJEROS GRATIS EN HOSPITALES PUBLICOS. @HORACIOLARRETA	@botedeltitanic1 @ernestoberni @horaciorlarreta Hay cosas buenas también en las plazas gimnasia para jubilados
@elcitizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitizen Yo pienso igual. Eso es lo han demostrado internacionalmente.
@elcitizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitizen Pues yo también así lo veo

Table 20: Examples of alignment-labeled data: Twitter (2)

Sentence 1	Sentence 2
@elcitizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitizen Triste pero yo tambien pienso lo mismo
@soyFranciscoALV: Primer publicación aquí	@soyFranciscoALV Muchas felicidades. Mucho éxito. saludos
RT @dominguezja: SATURNO es la canción más importante en la radio!! Felicidades!! @pabloalboran !! Te queremos!! https://t.co/1RWXUgdien	@dominguezja @pabloalboran Si, felicidades Pablete!!! Te lo mereces ♥
RT @PirryTv: A veces creo que todas estas traiciones entre políticos no son mas que puro teatro,después se sientan a tomar whiskey y a reír...	@PirryTv Yo no lo creo estoy seguro.
RT @PirryTv: A veces creo que todas estas traiciones entre políticos no son mas que puro teatro,después se sientan a tomar whiskey y a reír...	@PirryTv A veces? Jajajaja. Siempre!
Hermanos de Panamá y Honduras, @RusoEI23 lo dijo hace semanas y se cumplió... ¡Al Mundial y Repechaje! (...)	@FOXSportsMX @andremarinpuig @RusoEI23 @UpalabraMX Estoy muy contento por los hermanos Panameños! Esperemos que hagan un gran mundial!!!
@CARLOSLGUERRERO Y que dices de Alemania, ellos siguen vivos y Mejor que nunca, no es karma simplemente es dejar de hacer lo que se tenía que hacer.	@CARLOSLGUERRERO Bien lo dijiste mucho antes
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariciar, abrazar o dar palmas 🍌 https://t.co/kPw5QN2NBf	@AlejandroSanz Siempre uso las mías para tocarte la carita cada vez que te veo y gracias por devolverme siempre caricia con las tuyas. 🍌 te quiero. 💜
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio Suerte! solo hay que creer firmemente.
@soyFranciscoALV: Primer publicación aquí	@soyFranciscoALV Te deseo mucha suerte y acompañandote para siempre saludos
@soyFranciscoALV: Primer publicación aquí	@soyFranciscoALV Muchísima suerte para lo que viene
@caserocosas: QUE NINGUN ORGANISMO DE DDHH,HAYA PLANTADO REPUDIO TOTAL,POR LA MUE E DE CINCO ARGENTINOS,ES RARISIMO.	@caserocosas @lauale5 Los de los DDHH son como Joan M Serrat que todavía no salió a defender a los terro de Barcelona...
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de las Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky Ay, cómo me alegro. ¡Cómo me alegro! Enhorabuena y abrazo enorme.
RT @PirryTv: A veces creo que todas estas traiciones entre políticos no son mas que puro teatro,después se sientan a tomar whiskey y a reír...	@PirryTv De acuerdo, no lo dudes. Como crees entonces que todos ganan? Cada uno tiene su temporada y asi se cuadran!
RT @jpedrerol: No teniu vergonya (no tenéis vergüenza). https://t.co/HvC30fk9xr	@jpedrerol @kittlab No tenéis vergüenza.Y así nos podemos tirar toda la vida... ¿Es lo que quieres? Porque yo no.
@ehbildu: Tenemos que construir un muro democrático entre todos los demócratas @ArnaldoOtegi	@ehbildu @ArnaldoOtegi Ya hay muros así para demócratas como vosotros..
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Que así sea. Amén
RT @1Eme_: Sitios donde se reconoce la República Independiente de Cataluña#DuiARV #IndependenciaCatalunya https://t.co/P7IaZND5LA	@1Eme_ Si ya esperais en el primer dia que todos vayan a "Voy a reconocer la independencia de cataluña" asi como asi
RT @AnonymousVene10: Que memoria tan corta tenemos los Venezolanos cada quien siguió con su vida y olvidó a los q ya no tienen la bendición...	@AnonymousVene10 @untal_ro No es así, No manipules!!
@soyalannavarro: Algunos de los mejores momentos de la vida son los que no puedes contarle a nadie.	@soyalannavarro Así es es mejor guardarlo a uno mismo
RT @PirryTv: A veces creo que todas estas traiciones entre políticos no son mas que puro teatro,después se sientan a tomar whiskey y a reír...	@PirryTv Es correcto, llevan siglos haciendo lo mismo. Frente Nacional
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de escritorio, Colombia y Perú en el min88 tenían que jugar a clasificar a...	@fdbedout jajaja... En serio?! La verdad es que nadie entiende como en 3 minutos no arriesgaron la ida al mundial para clasificar a Chile
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Gracias por decir la verdad!! 🍌
@MonederoJC: ¿Cómo es posible que sin presentar ni una sola prueba y que el Tribunal de Cuentas diga que la financiación de Podemos está...	@MonederoJC Algo de verdad habrá o mucha. Al tiempo, es sólo cuestión de tiempo.
@oriolguellipug: Pues he estado estudiando el caso un poco y creo que sé lo que pasó ayer con la . Lo explico aquí	@oriolguellipug @guillemborrell Suena muy bien como explicación de una posible estrategia

Table 21: Examples of alignment-labeled data: Twitter (3)

Sentence 1	Sentence 2
@charlieredz: Será porque @VictorAndresGB es íntimo del Presidente de LAN? Tanto que se ausenta del Congreso para ir al matrimonio del h...	@charlieredz @VictorAndresGB Vaya vaya interesante, buscaré información; la opinión pública merece saber la verdad y saber quién es quién
RT @Schuma78: ERC en público: "Esto no va de independencia, va de democracia"ERC en privado: "Por la independencia nos pasamos la democrac...	@Schuma78 Jajajajaja que raro que digan esto los independentistas ...
RT @jmdelalamo: Hay dos tipos de países: los que no reconocen la independencia de Cataluña y los que no saben lo que es Cataluña.	@jmdelalamo Cuanta Ridiculez!! Váyanse a vivir a una Isla Desabitada! Y sean Independientes y Felices!! Si viven en 🇪🇸 son Españoles Y Punto!
RT @GramaticaReal: Atención: la palabra «solo» ya no lleva tilde, nunca.	@GramaticaReal @dennischwartz No estoy de acuerdo con esa regla
RT @marley_ok: Nació Mirko!!! 17:26 horas de acá, 51 cm, 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicidades te mereces lo mejor del mundo
@auronplay: Por fin me han puesto los 280 caracteres soy muy feliz en este momento quiero mandar un saludo a mi abuela a ver si un día v...	@auronplay Felicidades por tu cumpleaños y por los 280 caracteres.
RT @Blue_Leader_: #DuiARVRECONOCIMIENTO INTERNACIONAL CATALUÑA https://t.co/WYQBS1rk5W	@Blue_Leader_ Jaja nadie lo va reconocer!
RT @luciano_eltirri: Sin palabras!! @cuervotinelli ❤️ https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli Que emocionante!!!!
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IPHONES 8??? ¿Qué esperas? Correleeee! http...	@IamKatieAngel si!!! ya estoy participando ojalá ganarme uno eres la mejor te quiero
@YouNowLatino: - ¡Mañana a las 22:30h! Tenemos una cita con ellos @rfuertespuch y @BeatrizGH_17 directo en @YouNowLatino htt...	@YouNowLatino @rfuertespuch @BeatrizGH_17 Que guay. Mañana nos vemos. Que ganas.
@YouNowLatino: - ¡Mañana a las 22:30h! Tenemos una cita con ellos @rfuertespuch y @BeatrizGH_17 directo en @YouNowLatino htt...	@YouNowLatino @BeatrizGH_17 @rfuertespuch Que buena noticia
@NunkMasKK Hace años vi esos documentales sobre las cárceles de Japón, estamos años luz de allí...	@NunkMasKK Si, por supuesto. Aca los hacen las vagos de lo que son.
@nathcampost: Buenos días! Y feliz día a todas las Bautisters	@nathcampost Feliz día a ti también
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buen día cómo estás?
RT @marley_ok: Nació Mirko!!! 17:26 horas de acá, 51 cm, 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok que alegría por favor!! Felicitaciones les deseo lo mejor
@Soymarioruiz: Hoy te quiero más que siempre.	@Soymarioruiz Hoy te amo más que ayer.
@CristianSamXXX: Alguno quiere despertarla? Viene con premio	@CristianSamXXX yo quiero y también el premio
@JavierBeret: Todo lo que me está pasando me está preparando para lo que quiero ser.	@JavierBeret Y ojalá sea así
@Mishifucker: Vosotros también le sonreis a los perros por la calle, no?	@Mishifucker @EvaDRiobello No es una buena idea. No lo interpretan como una sonrisa, sino como un gesto hostil
@juanbedoya73: Mientras tenga vida y redes, me encargaré de recordarles, que en este país el gobierno pagó a militares por asesinar civi...	@juanbedoya73 @DeCurreaLugo Buena idea para un país sin memoria !!!
@MarioBautistaMB: Tu sonrisa te dará un semblante positivo que hará que la gente se sienta mejor a tu alrededor.	@MarioBautistaMB Tu sonrisa me motiva a hacer varias cosas
@Juandedios_P: Descansa preciosa	@Juandedios_P Igualmente amor
@luisvicenteleon: Ligar por un default de Venezuela y pensar que eso sacará al gobierno es una pendejada...al cuadrado.	@luisvicenteleon Venezuela no cae en default porque se ligue, se aupe, se quiera., igual que ligar a un no ganador...
@anaisbernal Es evidente lo que el abogado, la "manada" y el juez están pensando: Si es p... se lo merece, vamos a...	@anaisbernal Que tristeza y que rabia. Que se puede hacer?
@anaisbernal Que tristeza y que rabia. Que se puede hacer?	@anaisbernal Dime que esto no es verdad, por favor?

Table 22: Examples of alignment-labeled data: Twitter (4)

Sentence 1	Sentence 2
@unostuitsahi: Esto es chistoso y preocupante a la vez	@unostuitsahi @DanielSamperO Cada día nos sorprenden mas , nos reimos incansablemente pero tambien nos preocupan.
@bryanmouque: Todos queremos libertad.	@bryanmouque Lastimablemente no todos la tenemos.
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Jesucristo es el camino , la verdad y la vida. Amén
@geografosubjeti: En la España democrática no ha parado de haber presos políticos	@geografosubjeti Pero no es lo mismo, hay que tener en cuenta el hecho diferencial, ser catalán, esto los hace inmunes impunes
@cesarioncamilo: Yo estudié gratis, (con beca bicentenario + beca complementaria) y no sabes el compromiso que sentí estudiando; y ahora...	@cesarioncamilo @sebastianpinera No sabe nada de nada. Solo repite como loro los guiones que le prepara Goebbels.
@gabrielrufian: Nuestra lucha este 21D es por la restitución de las instituciones, gobierno y Presidente legítimo de Catalunya, Carles P...	@gabrielrufian Claro, claro amigo... pero tranquilícese.
@VaLuZenere: Es que no puedo parar de ver Stranger Things. Estamos todos igual no?	@VaLuZenere Si Vale te entiendo , esa serie esta increíble!
@ziberial: Casi me echan del bar de la risa	@ziberial Muy bueno. Felicitaciones al que lo hizo
@carlosmatareal: Lamentable fallecimiento para todos los venezolanos. Freddy Galavis se nos fué esta madrugada. Abrazo y buen viaje quer...	@carlosmatareal Lo siento mucho. Mis mas condolencias para las familias.lo recuerdo de la novela Juan la virgen.
@XavierGodas: "A los presuntos violadores de Sanfermines, la acusación les pide 22 años. A mi hijo, por una pelea de bar, le piden 50 añ...	@XavierGodas @lluís_llaç No hay justicia ni nadie que entienda todo esto. Como madre y abuela, lo siento mucho. Animos.
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes y @elchapuzas Bases sorteo en https...	@pccomponentes @elchapuzas Mucha suerte a todo el mundo y a ver si toca a tope
@clarincom: Chubut: impulsan un proyecto para que la Policía pueda demorar a encapuchados en protestas	@clarincom Me parece perfecto
@thegilbertsosa: Wow los que me amaban hace 10 horas, son los que más están escribiendo cosas feas ... pero pues está bien	@thegilbertsosa Está muy mal lo que hiciste.
RT @sebasantanders: Si Macri, Kuczynski y el presidente de Honduras fuesen de izquierda, la prensa ya estaría haciendo reportajes en extens...	@sebasantanders Estoy de acuerdo contigo, si hasta Maduro a salido de la agenda.
RT @PirryTv: A veces creo que todas estas traiciones entre políticos no son mas que puro teatro,después se sientan a tomar whiskey y a reír...	@PirryTv Con estos políticos colombianos, me arrepiento de haber nacido en este país sin oportunidades.
RT @AcunaMarcos17: Uno de los días más felices de mi vida! Vamos Argentina 🇦🇷 https://t.co/h3zCvXHHbl	@AcunaMarcos17 Felicitaciones Marcos. Sos el obrero de la selección Argentina de fútbol. Estamos orgullosos de vos.
@MarioBautistaMB: Tu sonrisa te dará un semblante positivo que hará que la gente se sienta mejor a tu alrededor.	@MarioBautistaMB Asi es como estar contigo con esta bella sonrisa que solo tan verla le robas millones de sonrisas...
@ChicasDelCable: Volvemos en Navidad. Pero que quede entre nosotras.	@ChicasDelCable Madremia de mi vida que no puedo más eh!Me estáis dando unas promos impresionantes!QUE GANAS!!!
@primavikinga: Hoy más que nunca me repito a mi misma que de todo se aprende... no hay errores en la vida... equivocarnos es vivir... ha...	@primavikinga Este mensaje, como cada uno de los tuyos, siempre me levanta el ánimo
@680carlos: Las Universidades huyen del pensamiento crítico. Los alumnos son clientes no estudiantes. Es el negocio de la mediocridad	@680carlos Totalmente de acuerdo
RT @ege171: El President no acata el cese impuesto por el gobierno de un país extranjero. Justo lo que se esperaba del presidente de mi país...	@ege171 @Somxserem Y nosotros tampoco.
RT @EPN: México no reconocerá la declaración unilateral de independencia de Cataluña. Hacemos votos por una solución política y pacífica.	@EPN Y a nosotros que nos importa si se quieren separar en españa. Creo que hay asuntos mas importantes que atender en nuestro pais.
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 Sí, claro.
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de las Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky No sabe cuánto me alegro. Enhorabuena, querida
@kevsho9: Estoy por dar un show en el Luna Park y no lo puedo creer es para un evento privado, por eso no avisé nada, pero es hermoso...	@kevsho9 Ay que genial, me alegro mucho mucho! Felicitaciones por todo

Table 23: Examples of non-alignment data: Textbook (1)

Sentence 1	Sentence 2
¡De nada!	¡Hasta otro día!
¡Genial!	¿Qué tal tu clase de español?
Azules, blancas, rojas y amarillas.	¡Sumi! ¡Qué grande es tu mochila! ¿Qué hay?
Soy ingeniero electrónico.	¿Cuántos idiomas habla?
Soy Pedro, el profesor de español.	¿Quiénes son ellos?
Muchas gracias, señora Pérez.	¿Qué tal el viaje?
Sí, son emocionantes.	Aparte del fútbol, ¿qué otras cosas te gustan?
¡Buenas tardes!	¿A qué hora sale el próximo tren a Sevilla?
¡Es mi poeta favorito!	¿Y por la noche qué hicieron?
Soy ecuatoriana y nací bailando ...	Para mí, estas serán mis últimas vacaciones de soltera, así que debo aprovecharlas antes de volver a Ecuador.
¡Espero que sea así!	Después, hay que ir rápidamente a una comisaría.
Pues a mí me parece que han hecho un buen trabajo.	No puedo creer que hayan prohibido las corridas de toros en Cataluña.
¿Y quién sale perjudicado?	Pues, los operadores de telefonía.
prefiero quedarme aquí.	¿Y qué escribes?
¡He caminado como nunca!	¿Qué hiciste el fin de semana?
Si tengo tiempo el próximo fin de semana, iré con Carlos.	Ya verás cómo te gustará.
Está bien que se quiera mantener en forma. Eso no significa nada. Yo solo te puedo aconsejar que hables con él para preguntarle directamente qué le pasa. La confianza es lo más importante en una relación.	Perdón por molestarte con mis problemas.
Es ideal. Dispone de un jardín hermoso con un lago romántico.	Parece que va a ser una celebración por todo lo alto ... No sabía que este evento tuviera tanta importancia en vuestra cultura.
Cambia sus zapatos planos por unos de tacón.	¡Qué interesante! ¿Y cuánto dinero se suele gastar en una celebración de este tipo?
Me alegro por ella ... me extrañaba que aún no estuviera casada. Es una chica espléndida, en la que se puede confiar.	¿Y lo vuestro por qué no funcionó?
En aquel entonces éramos unos críos ... tan solo teníamos 17 años.	Oye, la próxima vez tenemos que ir juntos.
El área geográfica de influencia de estos pueblos indígenas está distribuida en todo el país, específicamente por el occidente, el sur y la zona central.	¿Tienes datos sobre la población?
Pero tanto si vienen como si no, nosotros vamos, ¿eh?	Pues sí que te gusta ir a ese restaurante, no lo hubiera imaginado ...
Pues sí que te gusta ir a ese restaurante, no lo hubiera imaginado ...	No es eso, más que nada me apetece salir de casa ...
No es eso, más que nada me apetece salir de casa ...	¿Sabías que en una de las salas venden piezas de artesanía de Chile?
Pues no lo es. Que yo sepa, no tiene más de diez años.	Oye, antes de seguir hablando de otras cosas, más vale que llamemos para reservar una mesa.
Además, pueden elegir el horario que más les convenga.	Un momento, vamos a discutirlo y ahora volvemos.
¿Y si volvemos mañana?	Mira, lo del precio es lo de menos.
¿Y eso? Pensaba que no te gustaban los viajes sacrificados.	Necesito un respiro ... Tengo que ordenar mis ideas.
Necesito un respiro ... Tengo que ordenar mis ideas.	Pero, ¿no estabas pensando en buscar un trabajo?
Con la crisis, la situación se ha puesto muy mal.	¿Sabes que se ha convocado a un concurso de plazas en el Ministerio de Cultura?
Sí que te presionan ... Y yo que te dije que no tendrías ningún problema ...	Están deseosos de que me licencie y que busque un trabajo lo antes posible.
Bueno, aun nos queda una hora y media.	Ya sabes que ese restaurante siempre está lleno, espero que no nos falte tiempo para llegar a la hora.
De hecho, no conozco a nadie que lo haga.	Menuda excusa tienes para no reciclar. Millones de personas que piensan como tú agravan la contaminación atmosférica.
No te lo tomes a broma. Es un tema muy grave y alguien tiene que ponerle remedio.	Por el hecho de que yo recicle, el medio ambiente no va a mejorar.

Table 24: Examples of non-alignment data: Textbook (2)

Sentence 1	Sentence 2
Sí, aquí es.	Me inscribí por Internet. Mi nombre es Carolina Hernández.
Coreana. Soltera.	¿Dónde vive actualmente? Domicilio y teléfono.
Muy bien. Tome asiento y espere a que la llamen.	Aquí siempre preguntan el apellido materno, ¿te fijas?
Me preguntó que dónde y cuándo había nacido y mi estado civil. También me preguntó que dónde vivía.	Parecía enfadado.
No encuentro el formulario, ¿dónde lo puse? ¡Ah! Aquí estaba. Pensé que lo había puesto en mi cartera.	Sus pasaportes, por favor. ¿Son ciudadanas de la Unión Europea?
Muy bien. Todo está en orden. Feliz estancia y que disfruten en nuestro país.	Quisiera pasar al servicio, ¿en dónde nos vemos?
No, señor. Nada.	¿De dónde vienen?
Hemos encontrado una que tal vez sea la suya. Aquí viene.	¡He pasado un gran susto! Ahí traía mis documentos para inscribirme en la universidad.
Mejor pídele dos itinerarios, uno para mí también.	¿Billetes de ida y vuelta o sencillos?
Muchas gracias, señor.	¡Caramba! Tras cobrar tan caros los billetes todavía tuvimos que pagarle un seguro.
Gracias. Muy amable.	¡Señorita, olvida algo!
¡Me encantan los muñecos de peluche! Muchas gracias, Paco.	¿Soy el primero en llegar? ¿Dónde están todos?
Bertha Suárez. Mucho gusto.	Los dejo un momento. Voy a la cocina con los muchachos que están preparando unos canapés.
Los dejo un momento. Voy a la cocina con los muchachos que están preparando unos canapés.	Me puedes llamar Paco, mis amigos me dicen Paco. ¿Vamos a tomar una copa?
No, señor. Su licencia local para conducir es suficiente.	Disculpe, señorita, aquí en España cuando se saludan se dan dos besos, ¿verdad?
Sí, pero me daba un poco de problemas la batería.	Te felicito por el estreno. Entonces, nos veremos mañana. Chao.
¿Me permiten sus abrigos?	¡Qué lugar tan exótico! ¿No será demasiado caro?
No, no es tan caro. Espero que les guste.	Aquí tienen la carta y la lista de vinos. ¿Les apetece un aperitivo?
Está delicioso.	¿Sabías que mañana pasan el partido de fútbol entre Corea y Grecia en el Canal Plus?
¡Es Pak, lanza y ... goooool! ¡Era imparable!	¿Adónde quieres ir a cenar?
¿Sí? Entonces, haré como me sugiere y abriré una cuenta en dólares.	¿Quiere una cuenta corriente o de ahorros?
Claro que sí, y no le cobramos comisión, como le dije antes.	Por fin, mi esposo estará feliz con la cuenta en dólares, como él quería.
Por fin, mi esposo estará feliz con la cuenta en dólares, como él quería.	Aquí tiene su talonario y la tarjeta para el cajero automático.
Sí, le agradezco.	¿Funciona ahora?
Espere un momento, lo veré en seguida.	¿Cree que será muy difícil?
¿Cree que será muy difícil?	Hace un ruido extraño.
Creo que yo sí cumplo con todos los requisitos. Domino bien el inglés y el francés.	¿Tiene usted disponibilidad para viajar?
Sí, absolutamente.	Quien puede viajar tiene buenos puntos en la selección.
Gracias por su tiempo, hasta mañana.	Los que están afuera han presentado ya esta entrevista.
Le voy a poner un poco de anestesia.	Ahora dígame, ¿cómo se siente?
Me caí de una escalera.	¿Cuánto tiempo hace que le pusieron una inyección antitetánica?
No recuerdo exactamente.	Entonces, pregunte a quien le puso la vacuna. Luego me avisa.
No he podido ir a trabajar. El asistente que me ayuda se ha encargado de todo.	¿Hay alguna persona diabética en su familia?
El pescado que comí anteayer estaba muy fresco.	Vuelva mañana. Necesito hacerle unos análisis que complementen el diagnóstico.
Cuénteme cómo fueron los hechos.	¿Algún objeto desaparecido?

Table 25: Examples of non-alignment data: Twitter (1)

Sentence 1	Sentence 2
RT @Vitalisioo: RT si la ves gris y celesteMG si la ves rosa y blanca https://t.co/8HFeAsfWgv	@Vitalisioo Fin de la discusión xD https://t.co/lwJ8a2EC55
@Wereverwero Más que karma es porque ellos se parten rl lomo contra Brazil, Italia, Colombia, Argentina, etc. Y Mex... https://t.co/UboAaUmcBz	@Wereverwero Que hicieron ellos?
RT @SilvestreFDC: Al final me quedo con las mujeres q les gustó, no con las q me gustan! Triste terminar de hacer el amor y no tener de que...	@SilvestreFDC qué es eso?
RT @yFervazquez: Cri criminal 🍷	@yFervazquez Hoy se sale????
RT @mauriciomacri: No hay discusión. Messi deslumbra siempre. Aplausos para un jugador de otro mundo 🇦🇷🏆	@mauriciomacri @slgarelli_cnn Presidente compraron el partido dejense de joder
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/0g372ZKCWF	@pccomponentes @elchapuzas no quiero morir por favor
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/0g372ZKCWF	@pccomponentes @elchapuzas Quiero ser feliz
RT @Lhuchin: Chileno, no te olvides que el CSM que sacó a Bielsa fue Sebastián Piñerá.	@Lhuchin @lamesadelcafe 0 aporte el comentario
RT @Oficialmorris: Un PRI que se multiplicó y tuvo muchos mini PRIs https://t.co/9r8aFbKtbp	@Oficialmorris hay hijastros que son rebeldes
@CARLOSLGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!! ... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSLGUERRERO Contra los que México no puede en el proceso de osorio
RT @NoblexAr: Aclaremos desde ahora: La #PromoNoblex no era valida para Chile 🇨🇱 #Rusia2018 https://t.co/ZMKxC4La7B	@NoblexAr es más barato comprar un Tv...
RT @GianThurber: Los ticos son los mejores vecinos que la vida nos pudo dar, celebraron y alentaron a la sele como un panameño más.	@GianThurber jugaron a media máquina #ticos #puravida
RT @La12twittera: Una DEDICATORIA de los jugadores de @argentina para la PRENSA argentina https://t.co/0g372ZKCWF	@La12twittera @MiliCabj_5 @Argentina Cuanto le pagaron a Ecuador?
RT @ElMalaguero: POR FAVOR, LIO. https://t.co/FHCUpjhASR	@ElMalaguero Como me puedo bajar la genialidad?
RT @ZassCometro: El diario @elmundoes nos regala un artículo de la más depurada técnica de editorial cuñado y fuentes sin contrastar.Luego...	@ZassCometro @elmundoes @emergenciescat El Mundo nunca ha pasado de papel higiénico impreso.
RT @TomasCivetta: Que partido te perdiste Maldonado https://t.co/dRzxx5osZL	@TomasCivetta qué falta de respeto y empatía joder con la desaparición de una persona, cerebros de caca.
RT @JoseLChilavert_: @joaquinfalco La de tu esposa mejor, es mas rica	@JoseLChilavert_ @joaquinfalco de donde salió
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariciar, abrazar o dar palmas 🍷 https://t.co/kPw5QN2NBf	@AlejandroSanz En Costa Rica hay mucho desastre y damnificados por las inundaciones
@RuidoEnLaRed Pueden mostrar la factura y la declaración respectiva? Quien les dio acceso?	@RuidoEnLaRed En ese carro está el dinero del pueblo.
RT @lavozdegacia: 🇵🇪 El Valle de Arán anuncia un pleno el lunes para independizarse de Cataluña https://t.co/rxFd3ndt8 https://t.co/wUTvD...	@lavozdegacia voy a baqueira el puente de diciembre
RT @abarceloh25: Los senadores del PP jalean a las medidas del 155 mientras Rajoy las enumera. 🤔	@abarceloh25 @eulalia_fe No odiamos a España, tan solo pedimos que se nos respete...
RT @NomasKmail: Echarri no se fue del país.Brancatelli tampoco.Victor Hugo menos.Hebe no se suicidó.Promesas de campaña incumplidas:	@NomasKmail Querrán un incentivo?
RT @NomasKmail: Echarri no se fue del país.Brancatelli tampoco.Victor Hugo menos.Hebe no se suicidó.Promesas de campaña incumplidas:	@NomasKmail @BrunoRubnRuizD1 son amigas y alumnas de cristina
RT @cap0: Rajoy: "Violar las leyes, y por tanto la democracia, tiene consecuencias". ¡AY, QUE VA A DIMITIR!	@cap0 por un momento me he hecho ilusiones...
RT @Blue_Leader_: #DuiARVRECONOCIMIENTO INTERNACIONAL CATALUÑA https://t.co/WYQBS1rk5W	@Blue_Leader_ Ahora dónde jugará Piqué?

Table 26: Examples of non-alignment data: Twitter (2)

Sentence 1	Sentence 2
RT @Blue_Leader_: #DuiARVRECONOCIMIENTO INTERNACIONAL CATALUÑA https://t.co/WYQBS1rk5W	@Blue_Leader_ Quieren independencia, entonces que el Barcelona no juegue más en España.
RT @ilseCeroUno01: Llorando, estoy llorando os lo juro. https://t.co/R36NNIWxBJ	@ilseCeroUno01 @El_Formalito Cuando tus ideales te impiden ver la realidad. Y se puede aplicar a casi todos los aspectos de la vida
@resistenciadata Claro que sí!!!! Que oigan a esa gente que no está de acuerdo! 🇪🇸🇪🇸🇪🇸🇪🇸	@resistenciadata Puedo pedir un temita?
@resistenciadata Claro que sí!!!! Que oigan a esa gente que no está de acuerdo! 🇪🇸🇪🇸🇪🇸🇪🇸	@resistenciadata En manresa va sonar bien fuerte
RT @ensanro: La Declaración Unilateral Independencia es un error político que hará retroceder a #Catalunya y fortalecerá a la derecha española...	@ensanro @Caninator España es un Estado fallido.
RT @Elaguijon_: A ver lo que hace mañana este gilipollas sin DNI. https://t.co/st1LNM7nZ6	@Elaguijon_ ¡Adeu, 11€!
RT @edufeiok: El parlamento se limpia un poco https://t.co/vZ2sfD8QBd	@edufeiok @arielmnt1 Quien es el suplente de Devido?
RT @1Eme_: Sitios donde se reconoce la República Independiente de Cataluña #DuiARV #IndependenciaCatalunya https://t.co/P71aZND5LA	@1Eme_ Quien rie ultimo rie mejor
RT @Cumparini: Me topé a Kast saliendo de Santiago por el finde largo https://t.co/kNL7depNON	@Cumparini Pucha que son graves, era una broma
RT @luis_sisepuede: Enhorabuena, ya no tendréis que elegir entre la papeleta azul, la naranja y la roja para votar a la derecha #L6ARVCat ht...	@luis_sisepuede A la derecha se le tiene que poner las cosas fáciles, no sea que se lesionen pensando.
RT @kevsho9: VIDEO NUEVO! 🍷 Terminamos el mes de Halloween en el canal con un sketch y estoy acompañado 🤗 https://t.co/IRLPiMtDdf #Hallowee...	@kevsho9 Donde hay que poner el código?
RT @carnecrutaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrutaradio @emparmoliner Creo que a las próximas, el Borbón no se presenta.
RT @carnecrutaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrutaradio navidad está cerca
RT @Helioflores_mex: HOMBRES TRABAJANDO. Con el sudor de su frente y de sol a sol. https://t.co/a6hiilmZv2	@Helioflores_mex @TITOB_E_1 No merecen mi opinión, son plebes.
RT @carnecrutaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrutaradio @emparmoliner Está al caer.
RT @matiaslongoni: En septiembre se exportaron 20 mil toneladas de carne vacuna (31 mil toneladas peso res). Es el volumen mensual más alto...	@matiaslongoni Cambio de tema.
RT @renfrancis: Ya casi es Noviembre y yo sigo sin ti.	@renfrancis Vuelvan a MONTERREY
RT @CUBONoticiasMX: 🇲🇽 Mira @alfredodelmazo así "tratan" a los reos del Penal Neza Bordo ¿Algo qué decir? 🤔 https://t.co/uspCmVii2t	@CUBONoticiasMX @alfredodelmazo Firma la petición
@JavierMMoure @jinsua49 Con esa vocesita de nenita, hacé shhhhh y no jodas más. https://t.co/pRoROJKIXS	@JavierMMoure @ArielJob Leo todo lo que ponen de Sergio Maldonado y no veo críticas. Son simples apreciaciones!!!
RT @NicolasMDk: Cayó un aguaviva en la ventana del edificio. https://t.co/07OLB4yuWE	@NicolasMDk @PiNKeAdA Eso cuenta como tener sexo?
RT @NicolasMDk: Cayó un aguaviva en la ventana del edificio. https://t.co/07OLB4yuWE	@NicolasMDk @amitraviesas no respetan nada
RT @LaurenJauregui: Lo que quieras si me dices Por Favor https://t.co/4eMm4egUiiw	@LaurenJauregui @arianasharmoni vuelves a argentina?
@ziberial: Casi me echan del bar de la risa	@ziberial Ahora voy a dominar el mundo !!!
@DaniMateoAgain Dividir un país es un asunto muy grave con consecuencias, no debe ser tratado tan "emocionalmente" como hace él.	@DaniMateoAgain Hablan de fronteras y UE hay países que tienen menos habitantes que Málaga. Por ej. Luxemburgo.....
@_ju1_: Tras acabar una marcha españolista en Mataró. "Di viva España o te matamos". El agredido, hospitalizado. Difunde	@_ju1_ @soler_toni Donde estaba la policía?

Table 27: Examples of non-alignment data: Twitter (3)

Sentence 1	Sentence 2
@elena2760 Que tiempos aquellos	@elena2760 @adeladessire ¿Cómo se dejaron quitar eso?
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 Cual es el motivopor el cual deba ser gratis?...el cine es gratis?...el teatro es gratis?
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 No porque el fútbol es un derecho...
@xavieraldekoa Tocando el cielo con los dedos. Una fotografía maravillosa	@xavieraldekoa Mi padre viajero constante me dice que como la luz de África en ningún sitio.
@Samsung_Peru: Porque creemos en ti	@Samsung_Peru Ha sido siempre el capitán de selección peruano;
@estresfx_: dismenorrea, endometriosis, embarazo ectópico, ovario poliquístico....sigo?	@estresfx_ @Staptra No hay que ser feminista para saberlo. Deberían cerrar las cuentas.
@AlbanoDante76: Tertuliana de @laSextaTV diciendo que la vida en la cárcel es "amable" con "siestas de dos horas". Cuanta miseria.	@AlbanoDante76 @laSextaTV Imagine que estoy en Alemania y me pintan en mi tienda JUDÍO y una estrella de david.
@PatriotaNene: Argentinos que han venido a España y se permiten el lujo de querer trocear nuestra Nación. si quieres la deportación d...	@PatriotaNene Creo que la camisa tan apretada le impide un correcto riego sanguíneo al cerebro
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani no entendi
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani mañana es lunes
@Soymarioruiz: Buenos díaaas	@Soymarioruiz A esa hora yo andaba en el quinto sueño
@anabelenroy_tve: ¿Cuánto tenemos que aprender de ellos?... Qué delicia de video	@anabelenroy_tve Mi chihuahua es el ser más cariñoso del mundo
@anabelenroy_tve: ¿Cuánto tenemos que aprender de ellos?... Qué delicia de video	@anabelenroy_tve No quiero enfadar a nadie, pero no veo a ningún gatito capaz de hacer algo parecido.
@EvilAFM: Estaba esperando este momento	@EvilAFM comprarse una Xbox one para disfrutar de la vida
@bryanmouque: Me encantaría poder dormir.	@bryanmouque @Coders_Boys hoy es el cumple de mi sobrina y sobrino que cumplen 4 años desde Guadalajara
@PabloMM: 5 tiparracos que violaron a una cría de 18 años en un portal piden respeto a su intimidad.	@PabloMM ahora eres juez??
@8choOficial: Termina esta frase:Hoy hay ____	@8choOficial EL HUMANO ES RETRASO Y NO TIENE CURA PA E
@Rubi5: Como joder la saga entera de Pokemon en 30 segundos	@Rubi5 Alguien me explica que fue lo que dijo en inglés
@LSisco17: Sólo espero que la RAE acepte "Chavista" como sinónimo de mamagüevo.	@LSisco17 Es un acto amatorio no un insulto
@LSisco17: Sólo espero que la RAE acepte "Chavista" como sinónimo de mamagüevo.	@LSisco17 Se le dice directamente!!
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana.... :)	@ecosdelbalon @XabiAlonso El trabajo tiene su recompensa.
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana.... :)	@ecosdelbalon @XabiAlonso ¿Qué será lo próximo?
@tania_hermida: Ecuador pierde el rumbo. Michel Temer columnista invitado de nuestro diario público?@mfespinosaEC favor atender	@tania_hermida @mfespinosaEC Con tanto éxito económico, ¿por qué tiene a duras penas un 2% de aceptación popular?...
@srtabebi: No me ha bajao.	@srtabebi Se acabó la bebida
@tolerancia0: Piñera en sobre las "empresas zombis" y el debate con Mónica Rincón y Daniel Matamala	@tolerancia0 @camandov Arriba los corazones... y las manos.

Data ratio equalization As shown in Table 12, the ratio of the data was not equal (1:1). At the beginning, we tested with this data and noticed that the classifiers did not classify non-alignment pairs well. We assumed that the low rate of non-alignment pairs in corpus caused this low accuracy of non-alignment classification.

We tried to keep the data ratio of alignment and non-alignment close to 1:1. (The test data ratio was already set to 1:1; 70 alignment/70 non-alignment for textbook corpus, 100 alignment/100 non-alignment for twitter corpus.) For this purpose, we have implemented data augmentation with word dropout (see section 3.2.4). At this time, the word dropout rate was set to 0.1.

We had 980 alignment and 264 non-alignment pairs among 1,244 textbook pairs of training data. We octupled the 264 non-alignment pairs and made 2,112 pairs. The last 152 pairs were excluded, and 1,960 non-alignment pairs was acquired. We also doubled the 980 alignment pairs and made 1,960 pairs. 3,920 pairs were consisted of 1:1 ratio of 1,960 each. We use 392 pairs (10% of them) as valid data, and the rest 3,528 pairs as train data.

We had 1,226 alignment and 574 non-alignment pairs among 1,800 tweet pairs of training data. We quadrupled the 574 non-alignment pairs and made 2,296 pairs. The last 46 pairs were excluded, and 2,250 non-alignment pairs was acquired. We also doubled the 1,226 alignment pairs and made 2,452 pairs. By deleting last 202 pairs from 2,452 pairs, 2,250 pairs was acquired. 4,500 pairs were consisted of 1:1 ratio of 2,250 each. We use 450 pairs (10% of them) as valid data, and the rest 4,050 pairs as train data.

After adjusting the ratio to 1:1, the performance of deep learning model improved. The reason for improvement could be the data augmentation method itself or data ratio equalization. Identifying which was the main reason for the improvement was beyond the scope of this study, thus we did not investigated it deeply.

Table 28: Examples of non-alignment data: Twitter (4)

Sentence 1	Sentence 2
@ElJuanpaZurita: Bueno pues. Ahora sí. Noches noches noches noches.	@ElJuanpaZurita quiere irse
@itsandreix: El catalogo de ToyPlanet de juguetes si que sabe.	@itsandreix Precioso un poquito reconciliador con la raza humana
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio desagregado según sector socio económico?
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio acaso los pobres viven lo mismo que los ricos?
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio Eso significa vivir peor!!!
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio Cómo se llega a esa edad??
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio No es como lo mismo decir que el PIB per capita es el más elevado de latino América, no?
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio Agrégale a la curva el costo unitario del plan de salud...
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio entiendo que superamos por lejos en eficiencia al resto
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio EEUU tiene una esperanza de vida muy baja comparada con cualquier país
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio como se nota que duermes vendado...
@Mel_Lezcano: Mis victimas de siempre jajaja @MoliniersReal @MicaViconte	@Mel_Lezcano @MoliniersReal @MicaViconte No se como agradecertelas..
@Juanmalpr: Eduardo Inda pillado hoy con Granados antes de la comparecencia en el Senado.¿Grandes amigos? á	@Juanmalpr ¿Con quién están, lo véis??
@AlvaroUribeVel: Petro creó 50 mil contratos burocráticos en Bogotá, quebró la ciudad y aspira a la Presidencia	@AlvaroUribeVel Es hora de que nosotros el pueblo esclavizado busquemos la manera de salvar nuestro pais.
@Rubi5: Doblaje Español confirmed	@Rubi5 estoy en clase
@BangBangtan_Esp: En unas horas comienza "The late late show con James Corden" donde BTS se presentará.link para ver el show:	@BangBangtan_Esp @BTSxAMAs No quiero ir a la escuela
@BangBangtan_Esp: En unas horas comienza "The late late show con James Corden" donde BTS se presentará.link para ver el show:	@BangBangtan_Esp @BTSxAMAs unas dicen PM otras am cual creo
@Davanana: VAYANSE A LA VERGA	@Davanana @NiniaIndigo1 Que es esto, la filosofia de la generacion de gelatina?
@Davanana: VAYANSE A LA VERGA	@Davanana @ppmarino NINGUNA mujer le digan ERES HERMOSA la dañen.
@Davanana: VAYANSE A LA VERGA	@Davanana @ppmarino NO solamente hermosura EXTERIOR, Sino HERMOSA INTERIOR...
@acristofalo: Soy una buena alternativa para Benedetto, Sampaoli, teneme en cuenta.	@acristofalo no lo entendi
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @marcedeltiempo @FernandezAnibal En pocos meses será compañero de celda.
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @MariaLujan_Rey @FernandezAnibal Por que tiene celular ?
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @mis2centavos @FernandezAnibal Tiene internet el preso?
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @mis2centavos Dice @FernandezAnibal va a penetrar al arquitecto con este calor

Table 29: Augmented alignment and non-alignment data set

	Average length	Alignment	Non-alignment	Total	Ratio
Textbook	20.2628 (10.1314)	1,960	1,960	3,920	1:1
Twitter	28.3635 (14.1817)	2,250	2,250	4,500	1:1

Deep learning library-Tensorflow and pytorch Two types of deep learning libraries are mainly used: Tensorflow, developed by Google, and pytorch, developed by Facebook. Google’s tensorflow has been developed for research and production. On the contrary, Facebook uses pytorch for research and development, and another deep-running library caffe2 for production⁵. Because of this difference, Facebook’s pytorch code has been simplified compared to Google’s tensorflow code. The code is concise and has the advantage of being able to create prototypes quickly and easily, and is now preferred widely. In this study, pytorch was used to take advantage of its advantage.

Technique to mitigate performance fluctuations due to initialization Deep learning models change learning patterns according to the initial values of the network, and eventually the performance of the trained models changes. In order to alleviate this phenomenon, we experimented 5 times for each setting, taking the best value as the best performance that the model can produce.

BiLSTM As mentioned at subsection 2.2.3, the general LSTM considers the forward direction only, and it becomes difficult to reflect the information of the past. In order to mitigate this phenomenon, Bidirectional LSTM is considered which takes reverse direction into consideration (Schuster and Paliwal, 1997; Graves et al., 2013a).

⁵Reference site: http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture8.pdf

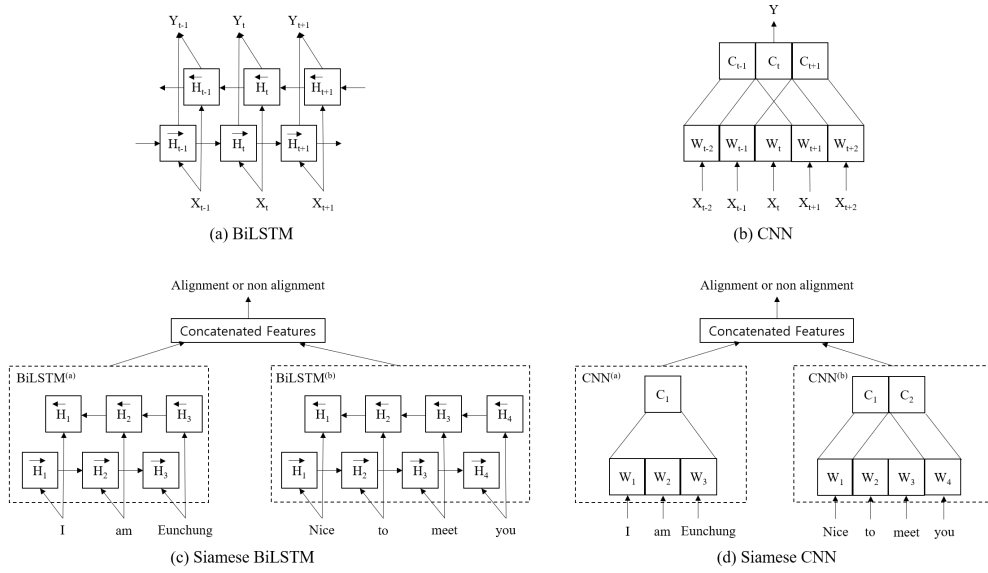


Fig. 26: Models used in the experiment

We experimented with this model⁶.

The major hyper parameters used in BiLSTM are as follows.

Table 30: List of Hyperparameters-BiLSTM

Hyper parameter	Value
Learning rate	0.001
Number of layers	2
Number of epoch	30
Batch size	150
Word embedding size	300
Hidden size	256
Dropout rate	0

CNN As mentioned at subsection 2.2.4, CNN moves the Kernel (window) to words in the sentence and changes the sentence into the feature. It is believed that it is pos-

⁶Reference code: <https://github.com/yunjey/pytorch-tutorial/blob/master/tutorials/02-intermediate/bidirectional-recurrent-neural-network/main.py>

sible to simultaneously consider surrounding words as much as the kernel size (Kim, 2014). We used $Kernel\ size = [3, 4, 5]$, $Kernel\ number = 256$ as in Kim (2014)⁷.

The major hyper parameters used in CNN are as follows.

Table 31: List of Hyperparameters-CNN

Hyper parameter	Value
Learning rate	0.001
Number of layers	1
Number of epoch	30
Batch size	150
Word embedding size	300
Kernel size	3,4,5
Kernel number	256
Dropout rate	0

Siamese-BiLSTM As mentioned at subsection 2.2.5, each sentence is put separately into BiLSTM, and the resulting features are grouped together. And we used it as an input data for logistic regression, and trained the alignment of pair sentences as an output⁸.

The major hyper parameters used in Siamese-BiLSTM are as follows.

Siamese-CNN This is the CNN version of the Siamese network(Hadsell et al., 2006; Yin et al., 2015). Siamese-CNN is difficult to use because NAN is floating even if setting is changed a little⁹ ¹⁰. After plenty times of tuning, we found the current setting¹¹.

⁷Reference code: <https://github.com/Shawn1993/cnn-text-classification-pytorch>

⁸Reference site: <https://github.com/harveyslash/Facial-Similarity-with-Siamese-Networks-in-Pytorch/blob/master/Siamese-networks-medium.ipynb>

⁹<https://discuss.pytorch.org/t/how-to-implement-siamese-network/2302>

¹⁰<https://discuss.pytorch.org/t/possible-data-parallel-memory-leak-for-siamese-network/4976>

¹¹Reference code: <https://github.com/yunjey/pytorch-tutorial/blob/master/tutorials/02-intermediate/convolutional-neural-network/main.py>

Table 32: List of Hyperparameters-Siamese-BiLSTM

Hyper parameter	Value
Learning rate	0.001
Number of layers	2
Number of epoch	30
Batch size	150
Word embedding size	300
Hidden size	256
Dropout rate	0

The major hyper parameters used in Siamese-CNN are as follows.

Table 33: List of Hyperparameters-Siamese-CNN

Hyper parameter	Value
Learning rate	0.001
Number of layers	2
Number of epoch	30
Batch size	150
Word embedding size	300
Kernel size	3
Kernel number	256
Dropout rate	0

3.3.3 Results

Fig. 27 and Table 34 show the result of alignment classification by each model. The result shows that Siamese CNN has the highest performance with the textbook data, and CNN has the highest performance with the twitter data. Deep learning models show superior performance than baseline, about 30-40% higher with textbook data, about 20-25% higher with twitter data. Notice that the result of Siamese CNN has even double accuracy rate than baseline. This suggests that deep learning approach is much better than marker-based (rule-based) approach.

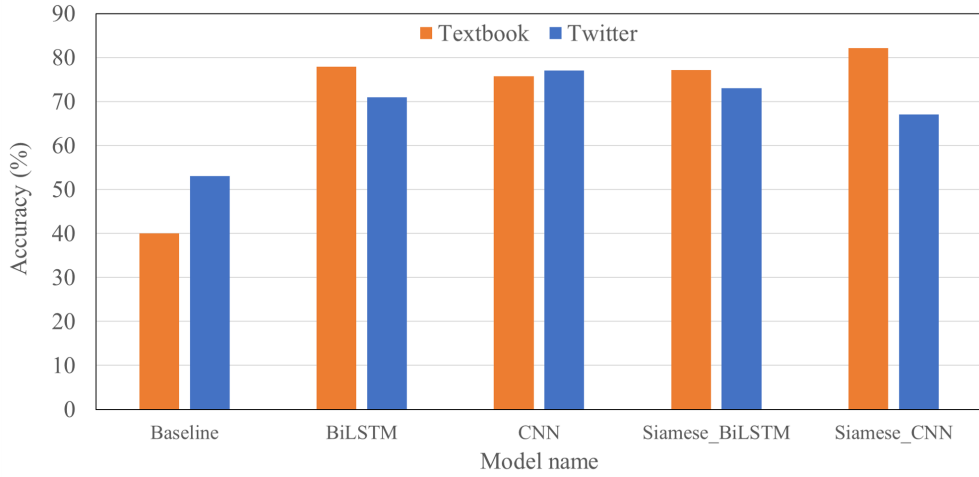


Fig. 27: Best results of models

Hyper-parameter tuning We chose best hyper-parameter setting for comparison of models in Table 34. To find best hyper-parameter setting, we performed numerous experiments. We modified hidden size $[32, 64, 128, 256, 512]$, and dropout rate $[0, 10\%, 20\%, 50\%]$. Total number of experiments is 800, $4 \text{ models} \times 2 \text{ data-types} \times 5 \text{ hidden sizes} \times 4 \text{ dropout rates} \times 5$ to mitigate performance fluctuations caused by initialization. The results are shown in the following graphs.

Table 34: Best results of models

Data type	Model	Accuracy (%)	Hidden size	Dropout (%)
Textbook	Baseline	40.00	Marker Lv.2	
	BiLSTM	80.00	512	0.5
	CNN	75.71	64	0.0
	Siamese BiLSTM	77.14	512	0.1
	Siamese CNN	82.14	64	0.0
Twitter	Baseline	53.00	Marker Lv.4	
	BiLSTM	71.00	512	0.2
	CNN	77.00	128	0.5
	Siamese BiLSTM	73.00	128	0.2
	Siamese CNN	67.00	512	0.2

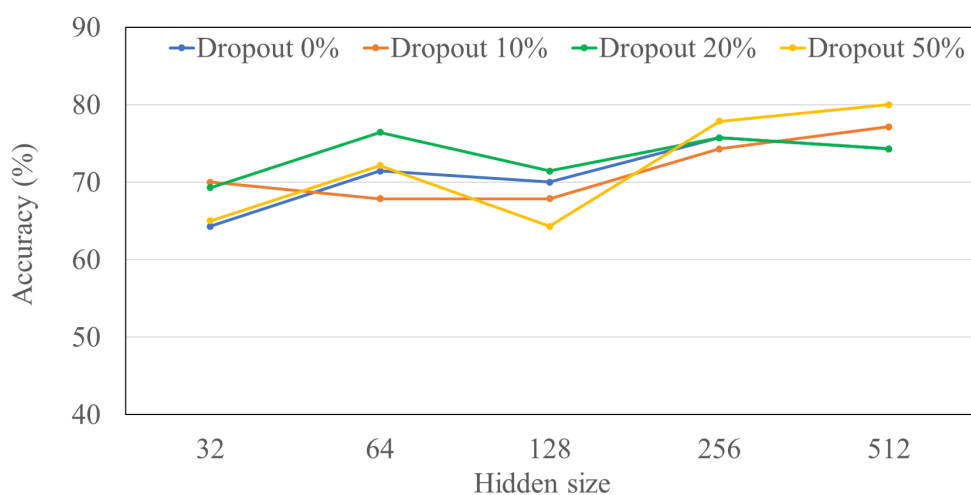


Fig. 28: BiLSTM result with textbook data. Accuracy for alignment classification as a function of hidden size

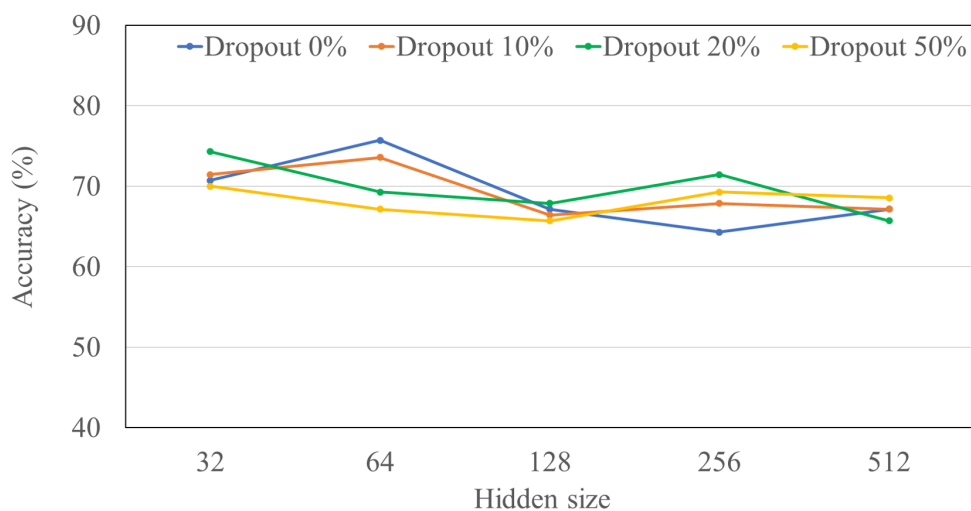


Fig. 29: CNN result with textbook data. Accuracy for alignment classification as a function of hidden size

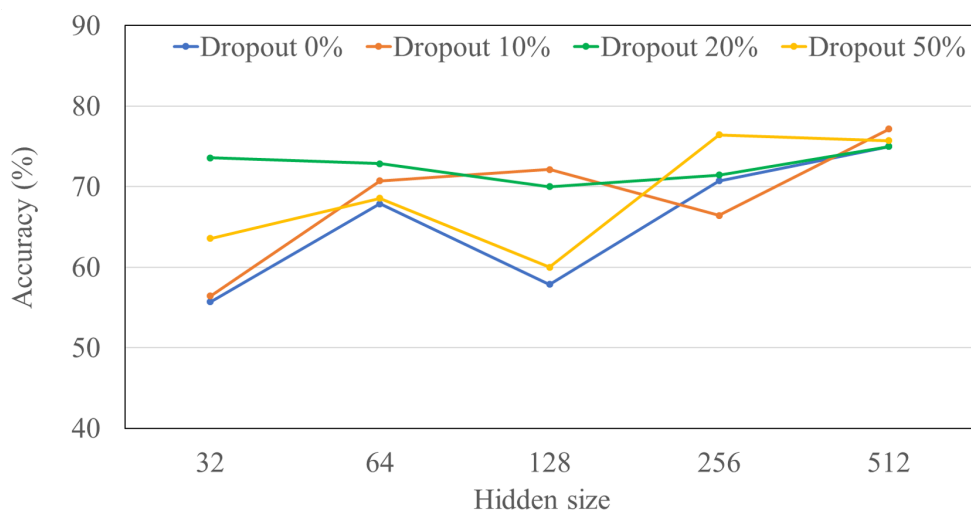


Fig. 30: Siamese BiLSTM result with textbook data. Accuracy for alignment classification as a function of hidden size

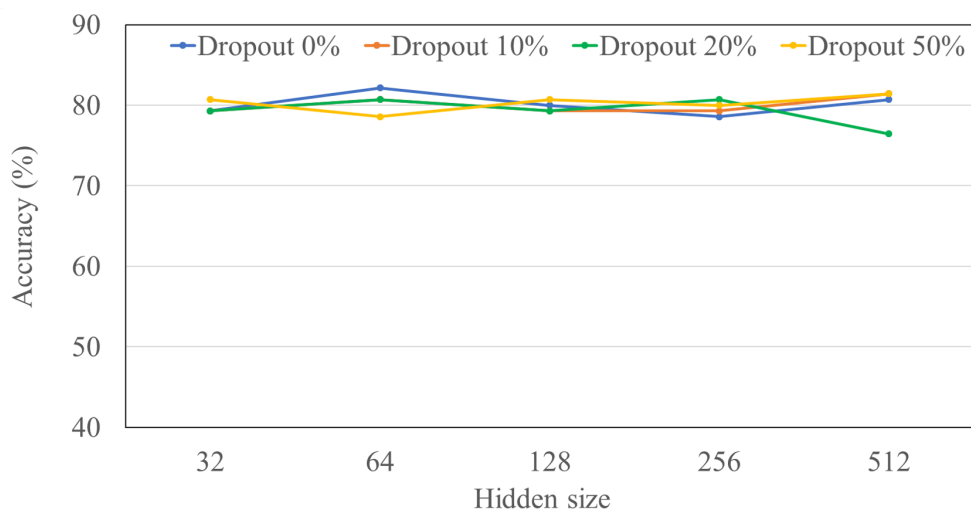


Fig. 31: Siamese CNN result with textbook data. Accuracy for alignment classification as a function of hidden size

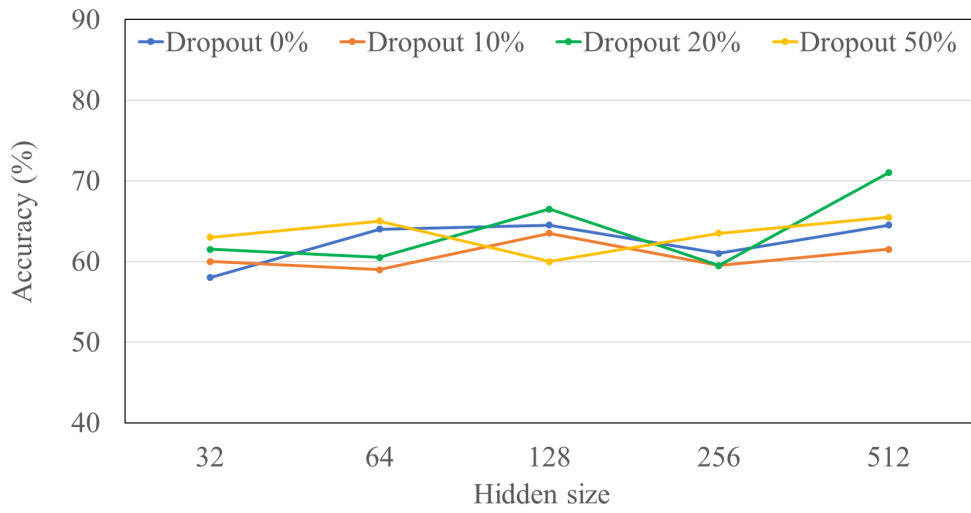


Fig. 32: BiLSTM result with twitter data. Accuracy for alignment classification as a function of hidden size

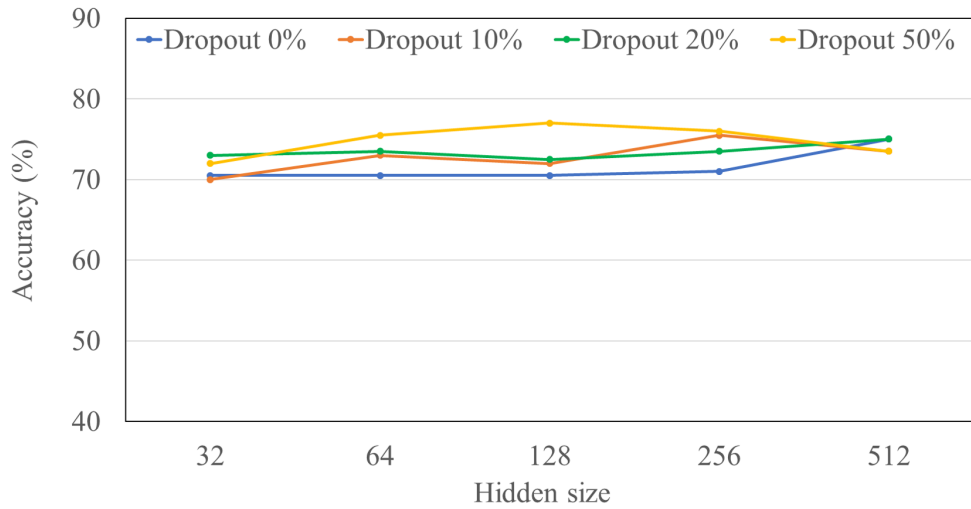


Fig. 33: CNN result with twitter data. Accuracy for alignment classification as a function of hidden size

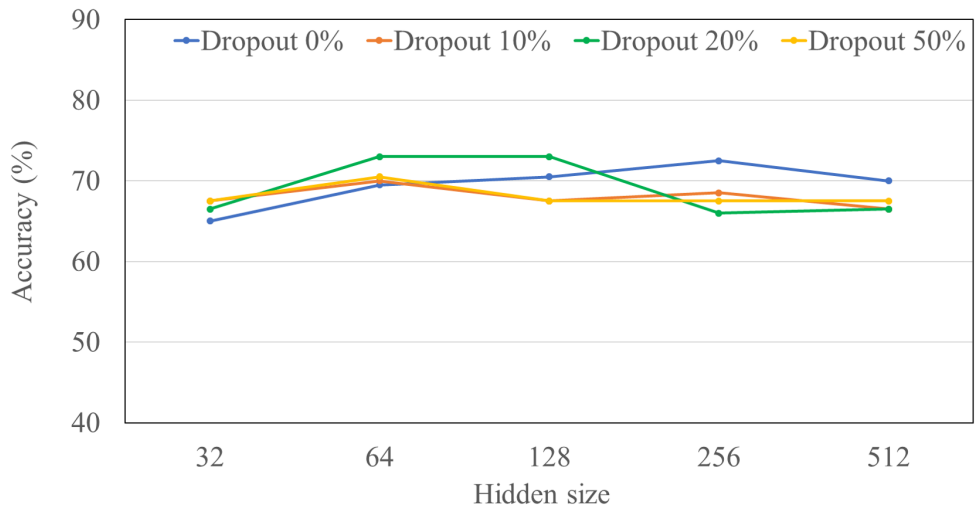


Fig. 34: Siamese BiLSTM result with twitter data. Accuracy for alignment classification as a function of hidden size

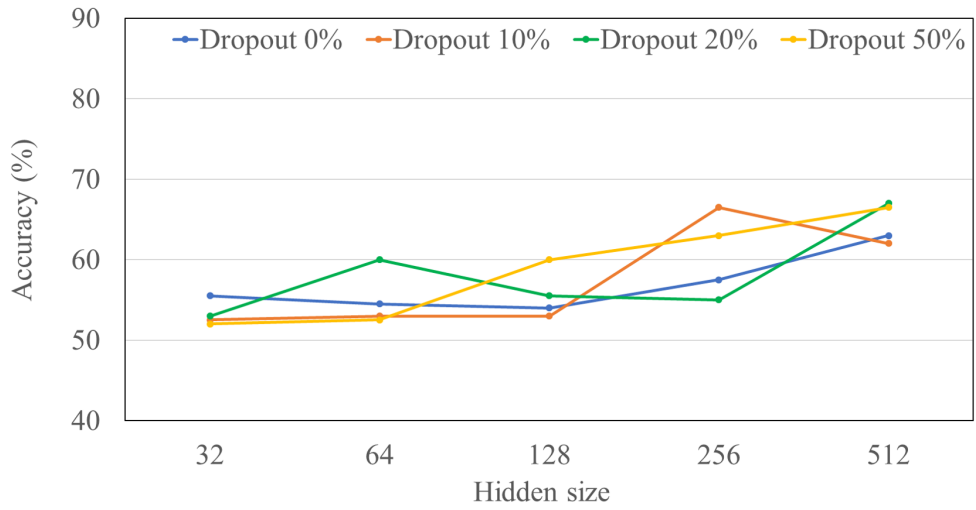


Fig. 35: Siamese CNN result with twitter data. Accuracy for alignment classification as a function of hidden size

3.3.4 Qualitative Evaluation

Since the results depend on the data, we must determine the model according to the data. We experimented a number of times to find the best hyper parameter settings. Each graphical representation of the results is shown from Fig.28 up to Fig. 35. The better we set up for our data, the higher the performance we may get, and the more appropriate classified data we will gather. As a result of experiments, it is recommended to change the type of model used in Classifier according to the nature of data. We have analyzed qualitatively the results of the highest performing Siamese CNN model in Textbook and the highest performing CNN model in Twitter. In order to evaluate the data qualitatively, the sentence must be determined semantically and appropriately.

In the textbook, Siamese CNN model, which showed the highest performance, was able to classify the following sentences as well. And an example of the most successful CNN model on Twitter is following together in Table 35.

Among the wrongly classified results, there were mixed sentences that were difficult to judge mechanically without lexicon or background knowledge.

From the perspective of pragmatic strategy in the context of maintaining the consistency of conversation, in the adjacent replies, the preceding words and concepts must be balanced by being co-referenced, reused or omitted, and the conversation proceeded in a way that introduces another word or concept. Therefore, in the alignment structure, the current or new information is balanced according to the speaker's speech.

Table 35: Evaluation of the results from the alignment classifier

source	model	Sentence 1	Sentence 2	judgment
Textbook	Siamese CNN (Test Accuracy of the model: 82.14%)	¿Y para beber? (And for drink?)	Para beber, una botella de agua mineral sin gas. (For drink, a bottle of mineral water without gas.)	Good (alignment)
		Muy bien. Todo está en orden. Feliz estancia y que disfruten en nuestro país. (Very good. Everything is in order. Happy stay and enjoy in our country.)	Quisiera pasar al servicio, ¿en dónde nos vemos? (I would like to go to the restroom, where do we meet?)	Good (non-alignment)
		¡Adiós! (Goodbye!)	¡Hasta la vista! (See you later!)	Bad (predicted: [N])
Twitter	CNN (Test Accuracy of the model: 77.00%)	¿Cómo habéis empezado la semana familia? (How have you started the family week?)	La he empezado increíble, gracias a tu me gusta soy la mas feliz del mundo ahora mismo. Gracias. (I've started it amazing, thanks to you. I like it. I'm the happiest in the world right now. Thank you.)	Good (alignment)
		Lo que quieras sí me dices, por favor. (Please tell me what you want.)	Vuelves a Argentina? (Do you go back to Argentina?)	Good (non-alignment)
		Cristo ha vencido la muerte: él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza. (Christ has conquered death: he is our resurrection and our life. Be witnesses of this message of hope.)	Es el camino, la verdad y la vida. Amén. (He is the way, the truth, and the life. Amen.)	Bad (predicted: [N])

3.4 Generation trained with labeled pairs

Our ultimate purpose is creating the system which generates aligned responses corresponding to input sentence. To create this system, we are going to use neural machine translation (NMT) model, which translate an input source language into a output target language. In our case, the input source language is *sentence 1* which is previous speech of the interlocutor, and the output target language is *sentence 2* which is aligned response.

Specifically using only the sentence pairs labeled as alignment among the 1,800 sentence pairs which used in section 3.3 as train data, we create an NMT model which generates *sentence 2* as aligned response when *sentence 1* is entered as an input sentence.

hypothesis By using the alignment corpus (1,226 alignment sentence pairs labeled by us in section 3.2), we would train the aligned response generation model (same with NMT model) and generate appropriate aligned response to the user.

3.4.1 Settings

As NMT model, Encoder-Decoder model (Cho et al., 2014; Sutskever et al., 2014) with attention are used (Bahdanau et al., 2014; Vinyals et al., 2015a; Luong et al., 2015; Xu et al., 2015) ¹² for this study.

The train and valid data are 1,100 alignment pairs / 126 alignment pairs, respectively. Test data is 100 alignment pairs from section 3.3.2.

The major hyper parameters used in encoder (BiLSTM), attention decoder (LSTM with attention) are as follows.

¹²Reference link: <http://pytorch.org/tutorials/intermediate/seq2seq-translation-tutorial.html>

Table 36: Hyperparameter lists of Encoder and Attention-decoder

Hyper parameter	Encoder	Attention-decoder
Learning rate	0.001	0.001
Number of layers	1	1
Number of epoch	50	50
Batch size	150	150
Word embedding size	300	300
Hidden size	256	256
Dropout rate	0	0
Max sequence length		50

3.4.2 Results

We experimented with encoder (BiLSTM)-attention decoder (LSTM). The generated result with test data input (= *sentence 1* of test data) is as shown in Table 37. Our goal is to generate a response which is similar to the target response corresponding to the input sentence. The result from the trained NMT model is the generated response.

3.4.3 Qualitative Evaluation

We make a lexicon with all the words from the data that we use at the training. When a word that does not appear in the vocabulary comes as input, the word is treated as [unk] token (unknown word).

After reviewing the results, as shown in the 2nd and 4th examples of Table 37, many [unk] tokens appear in input sentences. This would cause the trained NMT model to misrecognize the input sentence. As a result, the generated responses from the trained NMT model seem weird. In other words, as many [unk] token exists, recognizing input sentences would be hard and then misrecognizing would cause weird sentence. The 3rd example of Table 38 shows that when the main word in the input sentence is UNK, generates the weird sentence, bad case.

Table 37: Example of generated responses from the NMT model trained with labeled pairs

	Examples
Input sentence Target response Generated response	que dicen ?? ? [unk] todo te queda perfecto y dale con su puta “ trabajo “ el tachira a la muerte eterna claro que se puede subsidiar a la muerte eterna claro que hacer ...
Input sentence Target response Generated response	la justicia no es venganza ni [unk] . uno es inocente hasta [unk] [unk] . pero prisión preventiva es [unk] para [unk] ... tal cual . de acuerdo . así debe ser . y así es un show
Input sentence Target response Generated response	¿ cómo habéis [unk] la semana familia ? la he [unk] increíble , gracias a tu me gusta soy la mas feliz del mundo ahora mismo . gracias ... es muy estúpido
Input sentence Target response Generated response	[unk] la [unk] del [unk] a [unk] paz . me gustaría saber cómo hace para usar esta cuenta ud . desde donde esta . yo quisiera tocar las tuyas ..
Input sentence Target response Generated response	la fiscalía belga respondiendo a todas las preguntas ... igual que aquí !!! esto lo tenemos en cuenta ? es lo mismo con la gente de su país , me gustaría ... y no lo puedo entender !!!

Although no [unk] tokens are in the input sentences as shown in the 1st and 2nd examples of Table 38, there are cases where it is not generated well, bad cases. From the 1st and 2nd examples of Table 38, it was assumed that there was not enough train data to learn the appropriate generation.

We decided to increase the data to solve the [unk] problem and to generate better response. The method of increasing the data is presented in section 3.5, and further experiments are carried out after increasing.

Attention Matrix We conducted the inner test (which means investigate with train data) to confirm if attention matrix was well learned. As shown in Fig. 36, 37, the attention matrix was not well learned. Due to the nature of the task, the input sentence words and the target output sentence words may not match well. But more than this, the main reason would be the lack of train data to learn attention matrix.

We decided to increase the data to learn attention matrix well. The method of increasing the data is presented in section 3.5, and further experiments are carried out after increasing.

Table 38: Evaluation of generated responses from the NMT model trained with labeled pairs

	Examples
Input sentence <i>In English</i> Target response <i>In English</i> Generated response (Bad) <i>In English</i>	buenos días <i>good morning</i> hola buenos días <i>hello good morning</i> es que votar no hay nada más fiel a los catalanes y no se quede con la plata de mierda tienen la plata de mierda tienen una masa y duro con su lucha es un crack ! <i>Nothing is more faithful to Catalans than voting. People do not stay with the money. They have the money of shit. They have a mass, and have hard time with his fight. It's amazing!</i>
Input sentence <i>In English</i> Target response <i>In English</i> Generated response (Bad) <i>In English</i>	te amo <i>I love you</i> yo tambien te amo <i>I love you too</i> yo creo que calladita es un androide estoy seguro que no tiene cultura . . . <i>I believe that Calladita is an Android and I'm sure that it doesn't have culture ...</i>
Input sentence <i>In English</i> Target response <i>In English</i> Generated response (Bad) <i>In English</i>	españa no tiene [unk] posible <i>Spain does not have possible [unk]</i> a mi sin [unk] me encanta nuestro [unk] [unk] y nuestra [unk] de [unk] de casi todo . <i>to me without [unk] I love our [unk] [unk] and our [unk] of [unk] almost everything</i> y el instigador en el 2019 son los amo mucho que se puede <i>and the instigator in 2019 are I love you so much that you can</i>

Table 39: Generated responses for investigating attention matrix. Each sentence corresponds with Fig. 36 and Fig. 37

	Examples
Input sentence <i>In English</i>	Siempre se hacen bromas de la mala memoria del venezolano , pero jamás pensé que muchos se olvidarían de la lucha de ello ... <i>There are always jokes about the poor memory of the Venezuelan, but I never thought that many would forget about their struggle ...</i>
Generated response <i>In English</i>	que gran equipo ! <i>what a great team!</i>
Input sentence <i>In English</i>	De ustedes depende para qué utilizar sus manos ... acariciar , abrazar o dar palmas <i>It depends on you to use your hands ... caress, hug or clap</i>
Generated response <i>In English</i>	yo quisiera tocar las tuyas .. <i>I would like to touch yours ..</i>

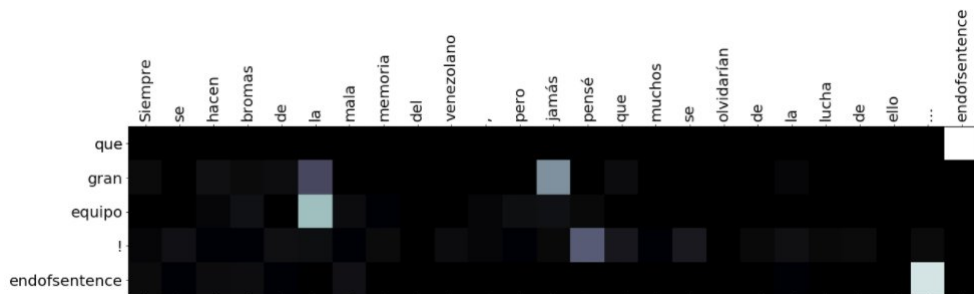


Fig. 36: Attention matrix from Table 39 1st example

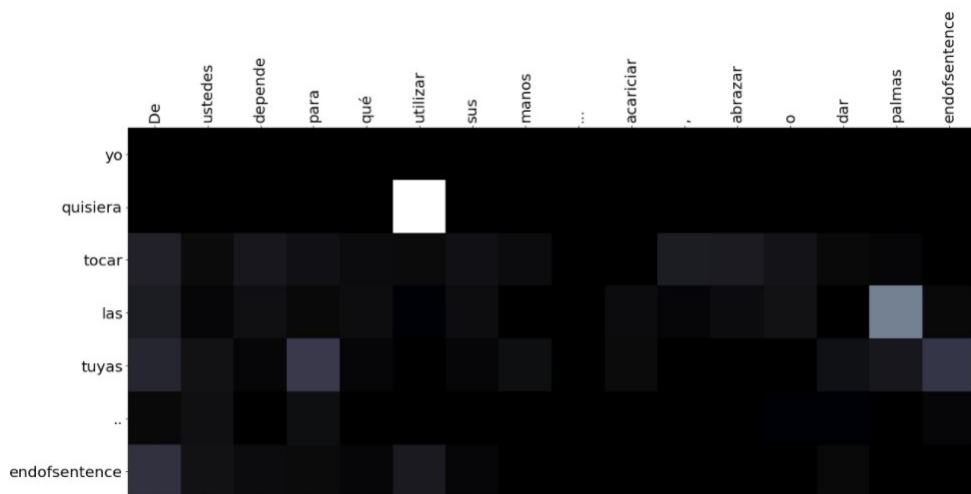


Fig. 37: Attention matrix from Table 39 2nd example

3.5 Generation trained with classified pairs

This section have the same purpose with section 3.4, and only one difference is the train data which is increased by following steps. To increase the data, we planed to use more twitter data and trained alignment classifier CNN, which perform 77% accuracy on twitter data (as shown in section 3.3.3). Detail as follows.

1. crawled more twitter data (50,000 sentence pairs).
2. utilized the alignment classifier CNN to classify a large number of sentences pairs (50,000 sentences pairs)
3. acquired the classified sentence pairs. 27,449 sentence pairs were classified as alignment pairs, and 22,551 sentence pairs were classified as non-alignment pairs.
4. used 27,449 alignment sentence pairs as train data for training aligned response generation model.

In other words, using only the sentence pairs classified as alignment among the 50,000 sentence pairs, we create an NMT model which generates *sentence 2* as aligned response when *sentence 1* is entered as an input sentence. Specifically, through the classifier, the 27,449 sentence pairs were classified as alignment pairs, and they are used for the aligned response generation model (See Fig. 38).

hypothesis 1 By using the larger alignment corpus (27,449 alignment sentence pairs classified by trained alignment classifier, CNN), we would train better aligned response generation model and generate more appropriate aligned response to the user.

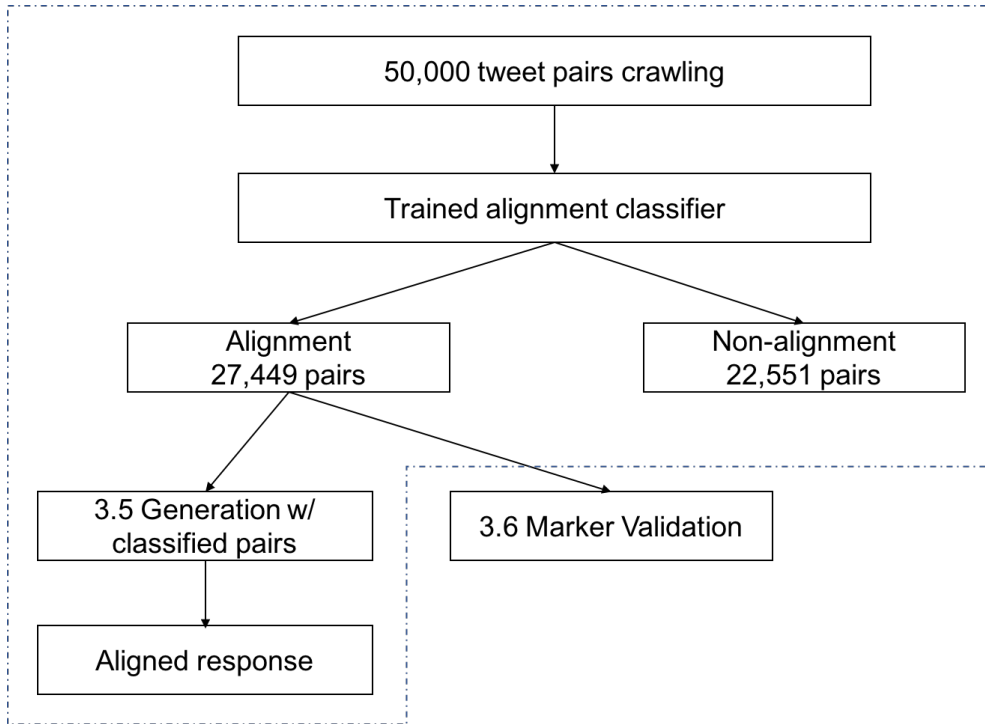


Fig. 38: Outline of the generation trained with classified pairs

hypothesis 2 Using large data would be better than using small data, even if large data would include misclassified data with 23% probability. Using CNN classifier which performs 77% accuracy would make over 23% misclassification. We expected that even though over 23% misclassified data was mixed to 27,449 classified alignment pairs, about 77% of 27,449 classified alignment pairs are rightly classified. The training would be weighted by many data, that is, 77% rightly classified data would lead the model to generate the aligned response. As a result, the aligned response generation model would be better.

3.5.1 Settings

The NMT model and the major hyper parameters used in encoder (BiLSTM), attention decoder (LSTM with attention) are same with section 3.4.1.

Only one difference with section 3.4.1 is the train data, changing from 1,226 labeled pairs that we directly labeled, to 27,449 classified pairs that the trained alignment classifier CNN classified. More specifically the train and valid data changed from 1,100 pairs / 126 pairs to 27,000 pairs / 449 pairs. Test data is same 100 alignment pairs from section 3.3.2.

3.5.2 Results

We experimented with encoder (BiLSTM)-attention decoder (LSTM). The generated result with test data input is as shown in Table 40. Notice that the number of [unk] token remarkably decreases in input sentence than the one from the model trained with labeled pairs (Table 37).

3.5.3 Qualitative Evaluation

Analysis of results In the examples of Table 41, the second example still produces a bad response, but the first and third examples generated the good-aligned responses. When comparing Table 37 and Table 40, the number of [unk] tokens is significantly reduced when recognizing input sentences. By contrast training with labeled data has vocabulary whose size 4,577, training with classified data has vocabulary whose size 35,925. As 31,000 more words occur in train data, the number of [unk] tokens reduced when test data comes as input sentence. As [unk] tokens are fewer in the input sentence, the trained model could recognize the input sentence well, and it would lead the trained model to generate aligned response well, a good case. But this

Table 40: Example of generated responses from the NMT model trained with classified pairs

	Examples
Input sentence Target response Generated response	que dicen ?? ? obviamente todo te queda perfecto qué pasa si no te vas a ver a mi le gusta la gana ... me voy a seguir rezando que me lo has dicho . es mi papa ...
Input sentence Target response Generated response	la justicia no es venganza ni morbo . uno es inocente hasta condena firme . pero prisión preventiva es válida para [unk] ... tal cual . de acuerdo . asi debe ser . son trabajadores ! ! !
Input sentence Target response Generated response	¿ cómo habéis empezado la semana familia ? la he empezado increíble , gracias a tu me gusta soy la mas feliz del mundo ahora mismo . gracias ... qué pasa si no te vas a decir esto ? por que no te digan , no me lo creo . se me ha hecho una falta ...
Input sentence Target response Generated response	agradecemos la visita del compañero a marcos paz . me gustaría saber cómo hace para usar esta cuenta ud . desde donde esta . qué pasa si eres un analfabeto una cosa que no te hace caso a las 4 días . si las hay que ver co ...
Input sentence Target response Generated response	la fiscalía belga respondiendo a todas las preguntas ... igual que aquí ! ! ! esto lo tenemos en cuenta ? es lo mismo con la gente de su país , me gustaría ... no mames , no importa que no le hagas caso a los demás cuando les conviene en sus ...

is only the case like the third example.

The first example changed from bad case to good case, not because of less [unk], but because the model could be trained with more various patterns as much as the amount of data increased. Larger data would lead the model to generate more appropriate aligned responses.

Table 41: Evaluation of generated responses from the NMT model trained with classified pairs

	Examples
Input sentence <i>In English</i> Target response <i>In English</i> Generated response (Good) <i>In English</i>	buenos días <i>good morning</i> hola buenos días <i>hello good morning</i> buenos días !!! que tengas una excelente día hecho por dios ! <i>Good Morning! Hope you have an excellent day</i>
Input sentence <i>In English</i> Target response <i>In English</i> Generated response (Bad) <i>In English</i>	te amo <i>I love you</i> yo tambien te amo <i>I love you too</i> que si sos hombre por que no te solo importa lo que importa es que mi jajajaja <i>if you are a man, then why don't you care</i> <i>what matters to me hahahaha</i>
Input sentence <i>In English</i> Target response <i>In English</i> Generated response (Good) <i>In English</i>	españa no tiene salvación posible <i>Spain does not have possible salvation</i> a mi sin embargo me encanta nuestro espíritu [unk] y nuestra capacidad de [unk] de casi todo . <i>to me however I love our spirit [unk] and</i> <i>our ability to [unk] of almost everything</i> me encanta el corazón ! no vale nada que a eso se lo ha dicho la verdad !!! <i>I love it, sweetheart! There is no value</i> <i>in telling the truth.</i>

Attention Matrix We conducted the inner test to confirm if attention matrix was well learned. Due to the nature of the task, the input sentence words and the target output sentence words may not match well. In spite of this, as the number of data increased, the attention matrix was learned better than Fig. 36, 37 as shown in Fig. 39, 40. Words like in Fig. 39 [con(with)-la(the)], [exc=excepto(without)-solo(only)] and in Fig. 40 [en(in)-tatuarse(to have a tattoo done)], [el(the)-primer(first)], [el(the)-tobillo(ankle)] were relatively well matched. These examples would indicate that the model still did not learn perfectly well but the model at least learned a certain amount. Due to learned attention matrix, the trained model generate better aligned response like the first and third examples of Table 41, good cases.

Table 42: Generated responses for investigating attention matrix. Each sentence corresponds with Fig. 39 and Fig. 40

	Examples
Input sentence	felicidades por alcanzar la meta en teletón para ayudar a compatriotas con capacidades especiales . exc ...
<i>In English</i>	<i>congratulations for reaching the telethon goal to help compatriots with special abilities. exc ...</i>
Generated response	solo eso faltaba que en su desesperacion se meta con la teleton . cuesta no ve más que la ...
<i>In English</i>	<i>That alone was missing in his despair to get with the teleton. The price does not seem more than the ...</i>
Input sentence	os tatuáis la fecha de nacimiento en números romanos , el nombre del perro en cirílico , el del hijo en japonés , “ te quiero ...
<i>In English</i>	<i>you tattoo the date of birth in Roman numerals the name of the dog in Cyrillic, that of the son in Japanese, I love you ...</i>
Generated response	yo aún sigo esperando a que salga el primer idiota en tatuarse el 155 , aunque sea el tobillo ...
<i>In English</i>	<i>I'm still waiting for the first idiot to come out to tattoo his 155, even if it's his ankle ...</i>

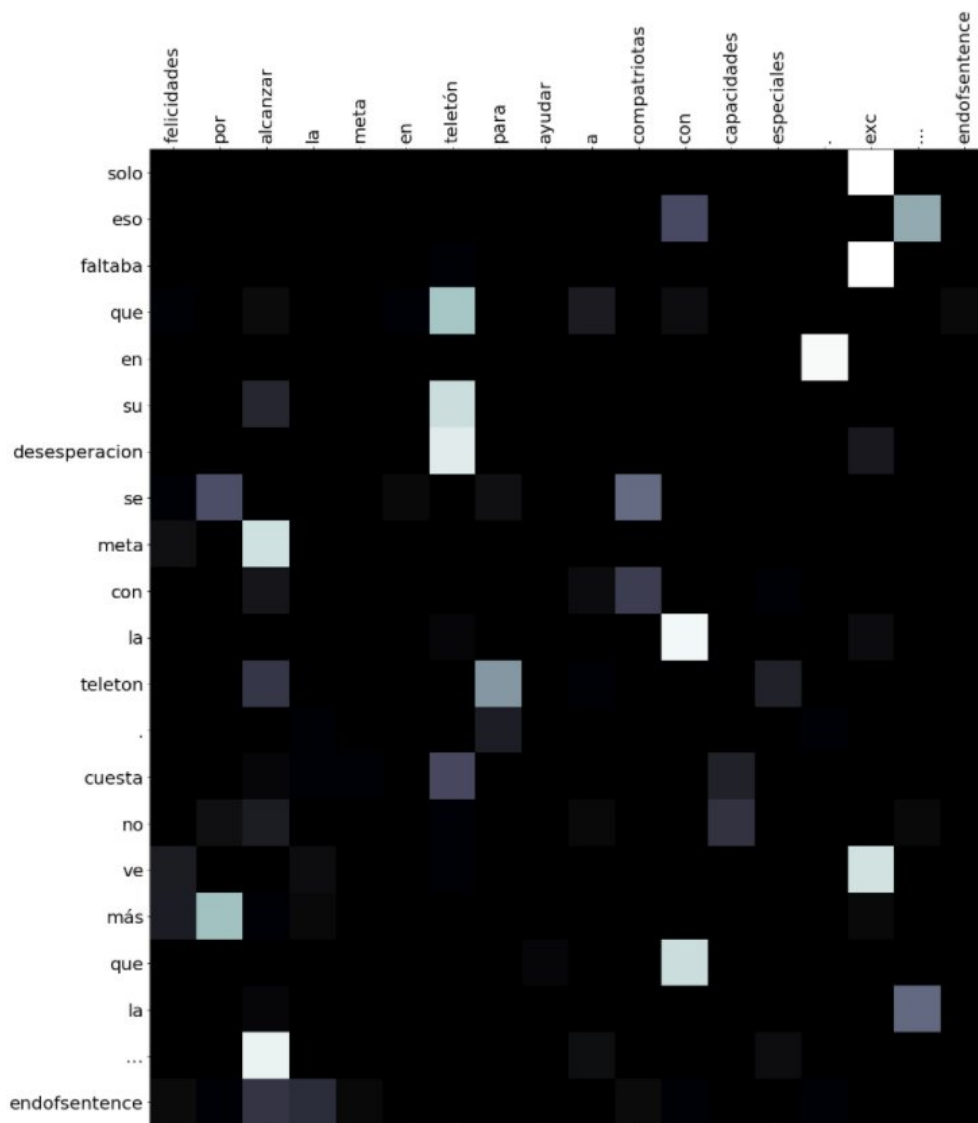


Fig. 39: Attention matrix from Table 42 1st example

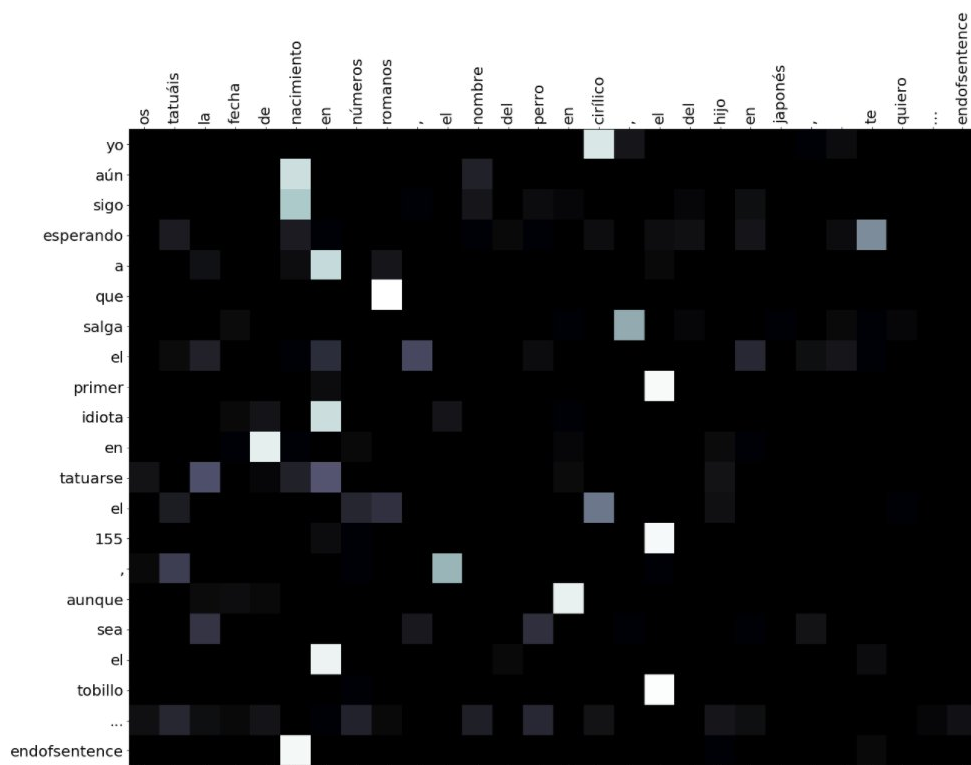


Fig. 40: Attention matrix from Table 42 2nd example

3.6 Marker Validation

In Spanish there are enriched lexical components expected as markers, and the alignment syntax may depend on such elements. We first made a list that could regard as marker in section 3.3.1. In addition, marker extraction was performed on the data classified by the trained classifier, CNN.

hypothesis 1 The top 50 frequent vocabularies in classified data could be alignment markers.

hypothesis 2 Among the top 50 frequent vocabularies in classified-alignment data and classified-non-alignment data, the different words can be markers.

3.6.1 Marker extraction from deep learning model

1. For the 50,000 unlabeled sentence pairs, apply the alignment classifier trained in Section 3.3 to distinguish the aligned response. Based on this, construct an alignment corpus.
2. Measure the frequency of each word in the alignment corpus¹³.
3. After deleting special characters/proper nouns, select the top 50 markers.

The frequency rate (total appearance / total replies in the corpus) of the words was measured.

¹³Reference site: <https://www.kaggle.com/anokas/data-analysis-xgboost-starter-0-35460-lb>

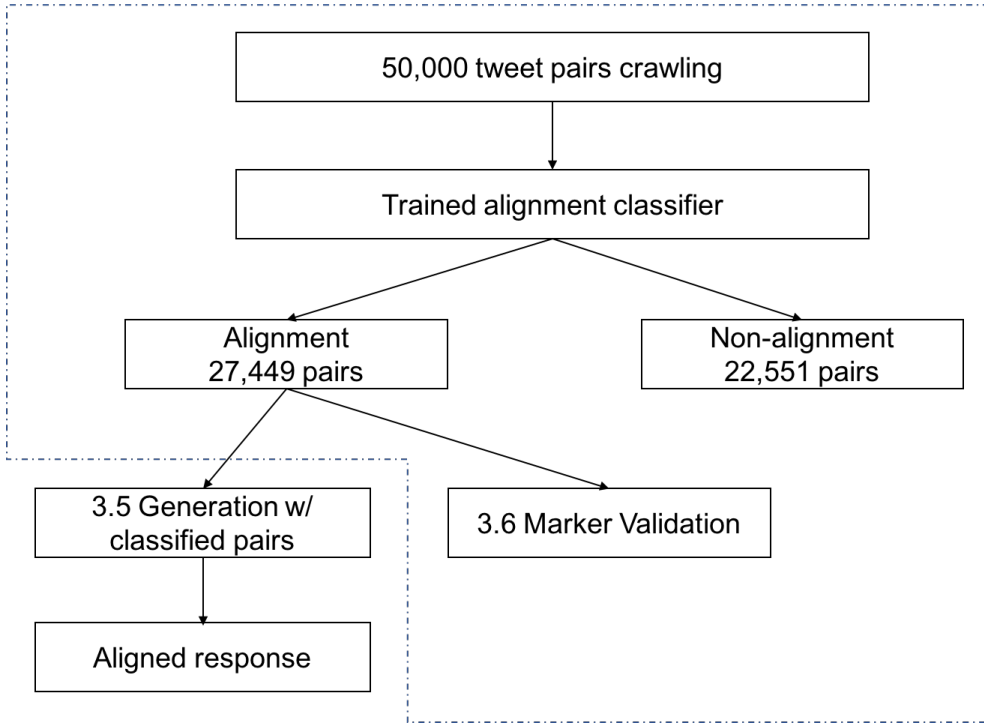


Fig. 41: Outline of the marker validation

3.6.2 Validation of Marker from human's and model's

Marker from human's The experimental results show low performance as shown at section 3.3, table 15. We predicted that human selected markers will guarantee better performance based on the theory, but it was different from our initial prediction. This is not only a motivation for research, but also a contribution of this study.

Marker from model's We classified the unlabeled 50,000 sentence pairs and extracted the markers by deep learning. However, the markers were not classified well, suggesting that deep learning does not classify linguistic data by a marker word, and the alignment should be categorized into other units than marker units.

RNN is known as a model that considers the order of the whole sentence, and

Table 43: Candidates for markers

Category		Examples
article	artículo	el, la, los, las, unos
certainty	certeza	siempre, jamás
conjunction	conjunción	pero, y, aunque
discrepancy	discrepancia	condicional; deber, hay que
exclusive	exclusivo	sin, excepto
inclusive	inclusivo	con, incluso
negation	negación	no, nunca
preposition	preposición	a, en, por, desde
pronoun	pronombre	lo, la, los, las, ti, ellos
quantifier	cuantificador	mucho, poco
tentative	tentativo	quizás, tal vez
supportive backchannel		sí, es verdad, de acuerdo, por supuesto, vale, entiendo, ¡Claro que sí!, ¡Qué buena idea!
Adverbial/prepositional phrase		así, ojalá, ahora (mismo), hoy, en seguida, contigo

CNN is known as a model that takes account the part (phrases) of the sentence. The fact that CNN performs better implies that it may be appropriate to pay attention to the phrase information for the alignment classification.

3.6.3 Qualitative Evaluation

The selected markers were manually categorized into the categories according to Syntactic and Conceptual as defined by Linguistic Inquiry and Word Count (LIWC) 2015 (Pennebaker et al., 2015). LIWC is a transparent text analysis program that counts words in psychologically meaningful categories. Empirical results using LIWC are known to be useful in demonstrating the ability to detect meaning in a variety of experimental settings, including attention, emotions, social relationships, ways of thinking, and individual differences.

We also excluded proper nouns, verbs, and stopwords from the analysis so that

we could focus more on linguistic features before analyzing markers obtained from actual language data and models.

We compared the high frequency vocabulary list presented by CREA (Corpus de Referencia del Español Actual) provided by Real Academia Española and selected the essential words for each category.

Table 44: Markers obtained from the model

Category		50 th top frequency from 50,000
article	artículo	la, el, los, las, unos, unas
certainty	certeza	siempre
conjunction	conjunción	que, y, pero, o, si, cuando, porque
exclusive	exclusivo	sin
inclusive	inclusivo	con, también . . .
negation	negación	no, nada
preposition	preposición	de, a, en, por, para
pronoun	pronombre	lo, se, te, me, su, le, tu, yo, este, esta, nos, eso, esto
quantifier	cuantificador	todo, mucho
backchannel		sí; mmmm, jajaja
adverbial/prepositional phrase		más, ya, muy, bien, mejor, solo, así

Effect of sentence classifier In the results, words appearing in both alignment (Table 45) and non-alignment (Table 46) are not considered to be alignment markers. The properties of the components are analyzed based on the results from the deep learning model, as described in the previous chapter, and extracted from the classified data set by its order of frequency. As we used the whole sentence, the unexpected *de* and *que* were ranked equally high in each experiment. This was a possible because we experimented with the entire sentences as a whole.

We expected pronouns, definite articles, or supportive backchannels to be at a high rate; however, thanks to the sentence classifier, unexpected interesting results were obtained. Among the top 50 words, there were 8 kinds of characters that ap-

Table 45: Alignment marker. Explicitly distinguished markers and high fluctuation of frequency rate (total appearance / total replies)

	Marked words	Absolute frequency	17	si	0.1939	34	todos	0.0777
1	de	0.9780	18	te	0.1896	35	son	0.0769
2	que	0.9774	19	para	0.1723	36	esto	0.0749
3	a	0.6922	20	del	0.1559	37	este	0.0728
4	la	0.6749	21	una	0.1464	38	qué	0.0687
5	y	0.6486	22	las	0.1372	39	hay	0.0687
6	no	0.5653	23	mi	0.1313	40	nos	0.0682
7	el	0.5029	24	pero	0.1290	41	tu	0.0673
8	en	0.4554	25	al	0.1109	42	o	0.0669
9	es	0.3901	26	más	0.1075	43	esta	0.0668
10	los	0.3325	27	le	0.1075	44	hoy	0.0648
11	por	0.3146	28	como	0.1043	45	muy	0.0605
12	lo	0.3077	29	ya	0.0984	46	ha	0.0602
13	un	0.2781	30	yo	0.0926	47	eso	0.0589
14	se	0.2565	31	todo	0.0914	48	ni	0.0588
15	con	0.2252	32	gracias	0.0901	49	porque	0.0575
16	me	0.1971	33	su	0.0851	50	cuando	0.0569

Table 46: Non-alignment marker. Explicitly distinguished markers and high fluctuation of frequency rate (total appearance / total replies)

	Marked words	Absolute frequency	17	una	0.1984	34	ya	0.0682
1	de	1.3294	18	las	0.1936	35	hoy	0.0609
2	la	0.9378	19	lo	0.1713	36	ser	0.0592
3	que	0.8528	20	al	0.1524	37	sus	0.0576
4	el	0.7710	21	me	0.1270	38	son	0.0539
5	a	0.7422	22	como	0.1181	39	o	0.0536
6	en	0.6246	23	su	0.1166	40	cuando	0.0529
7	y	0.6183	24	le	0.0954	41	si	0.0514
8	los	0.3851	25	mi	0.0897	42	todos	0.0504
9	no	0.3704	26	te	0.0830	43	porque	0.0496
10	un	0.3454	27	ha	0.0827	44	tiene	0.0469
11	es	0.2938	28	este	0.0818	45	sin	0.0449
12	se	0.2912	29	esta	0.0815	46	tu	0.0434
13	por	0.2524	30	más	0.0799	47	años	0.0426
14	con	0.2495	31	pero	0.0730	48	ahora	0.0406
15	para	0.2365	32	hay	0.0688	49	gente	0.0398
16	del	0.2235	33	está	0.0682	50	qué	0.0398

Table 47: Comparison of the eight vocabulary from alignment sentence pairs only

	Types	Accuracy (%)
		Twitter
Lv.1	Anaphors (pronouns, definite articles)	50.00
Lv.2	Positive/negative expression, Adverbs	52.50
Lv.3	Prepositions, Obligatory verbs	50.00
Lv.4	Conjunctions, Adverbs	53.00
Top 8 vocabularies only in aligned sentence pairs (<i>yo, todo, gracias, esto, nos, muy, eso, ni</i>)		61.50

peared only in alignment: *yo, todo, gracias, esto, nos, muy, eso*, and *ni*. Then, we reexamined vocabulary that appeared only in alignment, except for words in both cases. We attempted to retest these words using only 8 vocabularies. The re-test result was 61.5%, which was about 10% higher than the baseline result.

The word *yo* is the word for 'I' in English, with the highest frequency among the eight elements. Spanish can omit the subject because the verb is inflectional, but in real conversation, subject is often present to express the meaning of contrast or emphasis. And *todo* corresponds to 'every' but unlike English, it changes according to gender and number. Interestingly, the plural form of *todos* appeared in non-alignment. The word *gracias* is a noun to mean 'thanks'. It is mostly used idiomatically because it often expresses gratitude for previous speech of interlocutor in aligned responses. And *esto* is a pronoun corresponding to 'this', an abstract form that does not match gender or number. The word *nos* is used as a pronoun, a direct object 'us' or an indirect object. If someone asked a question to you (plural), the aligned response usually answer the question, using 'us'. And *muy* is an adverb corresponding to 'very'. In the answer, emotional expression or impression often appears. The word *eso* is a pronoun, a pronoun corresponding to 'that', similar to *esto* (this). This word is also a

Table 48: Eight examples with alignment markers

yo	A: México no reconocerá la declaración unilateral de independencia de Cataluña. Hacemos votos por una solución política y pacífica. B: Tú no, pero yo sí.
todo	A: Sí, es momento de cambiar un poco!! Te va a quedar divino! B: Todo te queda Hermoso a vos!
gracias	A: La Embajada de Chile es pensión acogedora para los demócratas perseguidos. Gratitud eterna. B: Gracias .
esto	A: Así ocultan los millonarios sus fortunas para eludir el pago de impuestos. B: Abramos más todavía los ojos. Esto es lo que hay que hacer.
nos	A: ¿Te indignan los que evaden impuestos mientras tú te matas a trabajar? Pues espera que Montoro les afine otra amnistía. B: Nos indignan quienes no pagan seguros sociales de sus trabajadores y luego dan clases de ética.
muy	A: Te amo. B: Te amo mucho más, me haces muy feliz con tu existencia
eso	A: Será porque @VictorAndresGB es íntimo del Presidente de LAN? Tanto que se ausenta del Congreso para ir al matrimonio... B: Eso pienso yo también. Vitocho cuida sus intereses ?
ni	A: Por muy poderoso que sea, hay algo que Donald Trump nunca podrá conseguir... Ser el Hermano Mayor de la Macarena. B: Ni el Muñidor de La Mortaja....!!!

basic form that does not match gender and number. It is used when it refers to abstract rather than to specific object. The word *ni* is the word for negation and is used in *ni ... ni* (neither ... nor). It is used in front of words to negate, such as nouns or verbs.

Despite the certain usage of these words, there is no helpful literature study to support theoretical explanations. It is difficult to capture the commonality of these words, except that they appear more frequently in the answers than in the questions. Although it is interesting to note that they only appeared in aligned responses and that the result was 10% higher than baseline when we experimented with these words. The deep learning, however, still performs much better than that.

Then, we examined the case of the most preposition *de*. In both cases, *de* appeared the most. This is because *de* is the most commonly used vocabulary in Spanish. Demonte and Soriano (2005) argued that *de* is a pure preposition and, obviously, not an evidentiality marker. The preposition *de* varies depending on the position or

contexts in the sentence:

(1) Posession

A: ¿**De** parte **de** quién?

B: **De** José.

(*Español básico* 1, Chpt. 10, p.177)

(2) Time

@GuillermoBotero El 7 **de** agosto **de** 2010

(3) Function word

a. Recepcionista: **De** nada. Lo que se le ofrezca, con gusto la ayudaremos.

(*Curso de Conversación y Redacción en español*, Chpt. 1, p.12)

b. Roberto: **De** acuerdo. Vamos.

(*Curso de Conversación y Redacción en español*, Chpt. 20, p.186)

Que is also the most commonly used vocabulary in Spanish. We should note that the Spanish *que* does not correspond to *that* of English. It can be misunderstood as a pronoun, however, it does not function like *that* as a pronoun (*eso*), but rather acts like *that* as a conjunction (*que*). They are not matched equally:

a. That's it. I knew that. That sounds interesting.

b. Eso es. Lo sabía. (Eso) suena interesante.

c. *Que es. *Yo sabía que. *Que suena interesante.

Here are the usage and examples of *que* in our data. Three main types are found:

(1) Relative pronoun

a. Roberto: Si quieres vamos a ese restaurante chileno **que** te guste tanto.

b. Natalia: ¡Genial! Es el restaurante **que** me gusta más que ningún otro.

(*Español intermedio*, Chpt. 14, p. 148)

(2) Comparative

a. RT @MonsieurSansFoy Aquí no hay nada más **que** mentirosos y embaucadores! A ver quien miente más para gatear al poder.

b. @soyunapringada_ mola, ya cobra más **que** la mayoría de los que suben zumos detox y demás mierdas.

(3) Conjunctions

a. Quiero **que** vengas/*vienes. [Subjunctive]

I want that you come-subj/*come-indic

“I want you to come.”

b. Es importante **que** tengas/*tienes cuidado [Subjunctive]

it is important that you have-subj/*have-indic care

“It is important that you be careful.”

c. Creo **que** *sea/es verdad. [Indicative]

I believe that it *is-subj/is-indic true

“I believe it is true.”

d. @PabloMM Es **que** esta noche ha sonado y aún no se ha despejado.

As in the above, *que* in Spanish mainly plays a role in connecting clauses. And in another case, as the proportion of Twitter data increases, the sentences which ignores accents and punctuation are included, that is *qué* is also mixed together.

(1) Interjection

a. Anabel: Menos mal que estás aquí. ¡**Qué** alegría encontrarte!

b. Lorena: ¡**Qué** sorpresa! ¿Qué haces aquí?

(*Español intermedio*, Chpt. 17, p. 177)

(2) Interrogative

a. Ruben: ¿Ah, sí? No lo sabía. ¿**Qué** tipo de voluntariado?

(*Español intermedio*, Chpt. 2, p.30)

b. @gabrielrufian Y por **qué** no fuiste a visitar al hospital a ninguno de los miles de apaleados? No te ha vergüenza?

The prepositions *de* also vary depending on the position or contexts in the sentence:

(1) Posession

A: ¿**De** parte **de** quién?

B: **De** José.

(*Español básico 1*, Chpt. 10, p.177)

(2) Time

@GuillermoBotero El 7 **de** agosto **de** 2010

(3) Function word

a. Recepcionista: **De** nada. Lo que se le ofrezca, con gusto la ayudaremos.

(*Curso de Conversación y Redacción en español*, Chpt. 1, p.12)

b. Roberto: **De** acuerdo. Vamos.

We firstly assumed in Spanish where are abundant lexical elements expected as markers characterizing the alignment structure. But the results were different from our earlier hypotheses. Even without specific marker, it can be an alignment structure, even if the sentence pair is not aligned, they may contain explicit markers as well. Then, in Spanish, it can be concluded that it is not possible to classify the alignment only based on certain markers. This is a contradictory result to what second language (L2) learners in general think about Spanish. The combination of sentence components is more important than specific rules that include markers. Focusing on some of the prominent vocabularies does not necessarily guarantee coherence.

Interestingly, despite the fact that the natural language sentence pairs were inserted just after the preprocessing, the sentence pairs were classified as aligned or non-aligned, and the performance was higher than markers. This can also be a way to study how deep learning works. The frequency of the marker could suggest the principle of operation. The deep learning model determines alignment structure based on a whole sentence level.

Interim summary We have obtained a model that can simulate about 80% of human judgment ability through several experiments. And the words which are frequently appeared in sentences are classified as markers. Except 8 examples, however, any significant difference was not found between the vocabulary found in the alignment and the vocabulary found in the non-alignment. While creating the correct answer data, we have confirmed that alignment criteria are not limited to specific words or vocabularies. This means that one or two specific components do not greatly influence the whole sentence. It is a meaningful study in that it was verified through

actual experiments based on empirical data.

3.7 Chapter summary

We collected Spanish textbooks and Twitter data, and trained alignment classifiers by adjusting the ratio of alignment / non-alignment data to 1:1 using data augmentation method. Then we performed experiments with baselines as selected markers, then compared the results with deep learning classifiers.

Here we also tried aligned response generation. To get more data, we utilized the best performance classifier, CNN, among the deep learning models that we designed in section 3.3. Since the best performance of the alignment classifier is 77%, it is expected that 77% probability would gather proper data and generate the alignment sentence with 77% probability. It did not work out well in every sentence, but we noticed some cases that properly generated the aligned responses. We analyzed some samples of generated results, and validated how aligned the sentences were.

In the marker validation section, we also utilized the best performance model, CNN, among the deep learning models that we designed in section 3.3. We compared baseline markers with those obtained from deep learning, and found better markers with deep learning models. It was, however, still better to classify sentence pairs with deep learning rather than marker dependent baseline.

Chapter 4

Discussion

Alignment is a strategy used by speakers to achieve communication objectives. We have developed a model that recognizes alignment to identify and validate markers, and tried to create a deep learning model that can make language research more efficient.

Ultimately, we aim to create a model that can generate responses that are close to each input answer. In addition, this study is not only for a linguistic model, but for establishing the standards and introducing deep learning to solve the problem which is difficult to define easily. The poor performance of the model reflects the difficulty of the task itself.

The goal is to create a system that always gives aligned replies to individuals. Regardless of any markers in the user's message, we want to create a system that provides context-sensitive responses at all times. Alignment research gives implications for the maintenance of conversational context. In a real-world conversation, it is important not only to maintain the dialogue context, but also to switch appropriately. It would be more user-friendly if the system could change the dialogue turns and topics moderately at the right moment.

4.1 Linguistic Feature of Spanish

Spanish is a group of Indo-European languages. Spanish is also called Castellano, originating from the Castilla region of northern Spain. Some of the Twitter we

collected for the study includes comments on Spanish in Portuguese, Italian, and Catalan. In this case, we exclude them because they are somewhat deviated from the actual research goal.

Each marker has some interesting properties. One of the main candidates for markers is the target pronoun, used in front of the verb, in order of [indirect object pronoun] - [direct object pronoun] - [verb]. In the case of third person, it is used as *lo*, *la*, *los*, *las* respectively according to the gender of the object word. Direct object pronouns are used in Spanish as a pronoun function in place of nouns that do not appear externally. And when the transitive verb forms a dual object phrase, the indirect object must appear in the form of dative pronoun (pronombres dativos) in the verb phrase.

Spanish objects Spanish has a very rich system of pronominal object clitics. Clitics are syntactically independent words or constituents, but depend phonologically on accents. Accusative clitics (*lo*) are direct objects, and dative clitics *les* and *se* are indirect objects. Both animate and inanimate objects are replaced by the accusative clitics *lo* and *la*, but only animate direct object NPs are preceded by what looks like the preposition *a*, which is typically assumed to be an animate marker or marked accusative case.

(1) a. ¿Ves a aquel chico?

Ø see-2ND-SG 3RD-PER-ACC that boy

“Do you see that boy?”

b. Sí, lo veo.

Yes, Ø him-ACC see

“Yes, I see him.” (*Español Básico I*, Chpt 9, p.163)

(2) a. María le quiere regalar su coche a su hermana.

Mary them-DAT want present her car to her sister

“Mary wants to give a present her car to her sister.”

b. Creo que sí. Se lo quiere regalar.

Ø think-1ST-SG that yes Ø her-DAT it-ACC-MAS want present

“I think so. She wants to give it to her.” (*Español Básico 1*, Chpt 9, p.166)

According to Torrego (1998), the distribution of the preposition *a* with animate direct objects appears to be sensitive to definiteness/specificity and lexical aspect of the verb and animacy/agentivity of the subject. For example, with stative and activity verbs, the preposition *a* is required depending on whether the subject is animate or not. If the subject is animate, then the preposition is required, but if the subject is inanimate, the preposition is not necessary. On the contrary, accomplishment and achievement predicates require the preposition *a* with animate direct objects, regardless of whether the subject is animate or not.

In Romance languages the syntactic feature of clitics has been an insistent issue. There has been a debate on whether clitics are XPs (or phrases) base-generated in argument positions, which then move to preverbal position (Kayne, 1975; Strozer, 1976; Rivas, 1977; Borer, 1984), or whether they are Xs (i.e. heads) that head their own agreement projections and act as morphological (agreement) affixes (Franco, 1993; Sportiche, 1996). While the essential feature of clitics is not particularly relevant for the purposes of this article, Franco (1993) and Sportiche (1996) assume that direct and indirect object clitics in Spanish are generated in two functional projections above VP and below TP – AgrOP and AgrIOP. Direct and indirect object DPs move to the Specifier of these functional categories and check accusative and dative case, respectively.

Since English lacks object clitics and has strong pronouns, the functional projections for clitics may be assumed not to be instantiated. (Alternatively, these projections may be assumed not to have weak agreement features.) In addition to syntax, semantics plays a role in the realization of objects and their corresponding clitics in Spanish.

Indirect objects are always preceded by the dative preposition *a*, and replaced by the dative clitic *le/les*. Clitic doubling, the co-occurrence of the clitic and the PP, is a possible option in all Spanish dialects. The clitic of indirect objects is generated as the head of functional category AgrIOP. The dative PP captures structural syntactic dative case through Spec-head agreement.

(3) a. ¿A quién le regalas estas flores tan bonitas?

to whom him/her-DAT present-2ND-SG these flowers such pretty

“To whom do you give these such pretty flowers?”

b. Se las regalo a mi mujer.

him/her-DAT them-ACC present-1ST-SG to my wife

“I gave them to my wife.” (*Español Básico I*, Chpt 9, p.166)

By contrast, dative clitic doubling is obligatory in some Spanish constructions. Those clitics are not related to structural cases, but appears to have semantic functions, related to affectedness or aspect.

The syntactic, pragmatic and semantic properties of subjects and objects in Spanish are shown:

In Spanish, clitic pronouns are commonly used as pronouns in place of nouns. However, when a direct object or an indirect object containing definite qualities is transposed by a verb, the accusative pronoun must appear and be used in a dual form.

Table 49: Feature specification of subject/object constructions in Spanish and English (Montrul 2004:130)

	Features	Subjects	Direct Objects	Indirect Objects
Spanish	-interpretable (strong)	Agreement	Structural accusative	Structural dative
	+ interpretable	Topic/focus	Marked (inherent) accusative	Marked (inherent) dative
English	-interpretable (weak)	agreement	Structural accusative	Structural dative

However, in addition to the pronoun usage, there are phrases that are mandatory in dual form, even though direct - purpose bore or indirect bore which appear in the same direct relation with oneself appears externally. This is the case when an accusative bore containing definite qualities is transposed to the verb.

If we look at this object transposition phrase, if the noun phrase mentioned in the transposition of the direct object is confined to the verb, it must be accompanied by a direct object pronoun. On the other hand, when a displaced noun phrase lacks definiteness, it can not be used like a direct object pronoun. The use of direct accusative pronouns in a dual form is associated with information related to information related to information (Vivanco, 2013; Groppi, 2008; Alconchel, 2002). In Spanish, the accusative pronoun is used in a dual form when the objective bore is a definite indication element of the speaker, and because the element having the definite qualities contains the old information qualities, the result is that the direct pronoun is used by the pseudo-. In other words, the reason why the dual form syntax with accusative pronoun is used is not the object pronoun function but the accusative predicate.

Null subject language Spanish has rich verbal inflection, which allows person and number information about the subject to be recoverable.

(4) a. ¿ Ø Conoces a alguien en este hospital?

know-2ND-SG 3RD-PER-ACC somebody in this hospital

“Do you know anybody in this hospital?”

b. Ø No, no conozco a nadie.

No, no know-1ST-SG 3RD-PERSON-ACC nobody

“No, I don’t know anybody.” (*Español Básico 2*, Chpt 3, p.47)

Toribio (2000) explains that Standard Spanish has strong nominal (PER, NUM) features in T and strong Nfeatures in Agr. By contrast, null subjects in English are typically not possible, except for imperative sentences. Therefore, English has weak nominal (also verbal) features of Agr and T syntactically, and subjects are realized overtly.

Table 50: The syntactic differences between Spanish and English with respect to the Null Subject Parameter (Toribio, 2000)

	N-features in T	N-features in Agr	Subject expression
English	Strong	Weak	Overt
Standard Spanish	Weak	Strong	Null

Zagona (2002), however, explains it is not the case that null and overt subjects are entirely optional. In fact, there are constructions that strictly require the use of null subjects, such as existentials with haber, weather verbs, and pleonastic constructions.

(5) a. Ø Hace buen tiempo. [weather verb] (*Español Básico 1*, Chpt 8, p.141)

b. Ø Hace años que no sé nada de ella. [time verb] (*Español intermedio*, Chpt 5, p.62)

c. Ø No hay ningún inconveniente. [existential] (*Curso de conversación y redacción en español*, Chpt 19, p.178)

d. Siendo entrevistador Ø es lógico que tuviera que ser objetivo. [pleonastic]
(*Español intermedio*, Chpt 4, p.52)

4.2 Advantage of Deep Learning

Recently, deep learning methods achieve the state-of-the-art performance in various machine learning tasks. It is known to perform better than rule-based approaches. Since feature engineering is not (or pretty less) performed in deep learning, deep learning requires pretty less cost. The deep learning method will learn the rules by itself from input and output data. It is a method by recognizing a pattern rather than specifying each rule.

The process used in this study indicates that deep learning could be helpful in theoretical linguistic study. Since the lexical components are very well developed in Spanish, we thought that there would be significant markers in the aligned responses. From our experiments and comparison with deep learning methods, however, there were no markers that showed a significant difference. That is, we found that a single marker is not very important in Spanish, which contradicts to our original expectation, as discussed in section 3.6. With deep learning method, which does not need any rules, we could have achieved this conclusion.

Chapter 5

Conclusion

We hypothesized that certain verbal elements quantify one's conversational style corresponding to categories of alignment markers. Aligned response refers to a conversational structure that continues communication in the context of the preceding utterance. It affirms or rejects the utterance of the interlocutor, or repeats the preceding statement. On the other hand, non-alignment refers to the cases in which the previous-mentioned topics are changed or the flow of the conversation is interrupted, or the words are not related to each other at all. Since the purpose of a general dialogue is to exchange information, it is common to be presented sequentially.

In section 3.3, the alignment classifiers were trained with 3,920 pairs in textbook corpus and 4,500 pairs in twitter corpus, after data augmentation. The performance of the classifiers was compared to the baseline marker experiments. The classifiers shows quite good performance (70 – 80%), better than the baseline (35 – 55%). Deep learning models are better to classify sentence pairs than marker dependent baseline.

In section 3.4 and 3.5, we also tried aligned response generation. The responses did not align well corresponding to every sentence, but we noticed some cases where the aligned responses were generated appropriately.

We expected that there would be significant markers affecting the aligned responses, because the lexical components are very well developed in Spanish. However, aligned response is not strongly influenced by lexical factors, and there were no markers that showed a critical difference, according to our experiments. With deep learning and marker-based approach, we found that a single marker is not very im-

portant in Spanish, which was not what we originally expected as discussed in section 3.6. If we had not used the deep learning method, we would not have achieved this conclusion. This process also indicates that deep learning could be helpful in theoretical linguistic study.

This study established the standards of linguistic alignment and introduced deep learning to classify alignment sentence pairs and to generate aligned responses. We have accomplished our purpose through various experiments. It is valuable as a research to focus on alignment problem in conversation, establish theoretical basis, and perform marker validation using deep learning. We are expecting to extend this methodology to other languages.

References

- Auli, M., Galley, M., Quirk, C., Zweig, G., 2013. Joint language and translation modeling with recurrent neural networks.
- Backstrom, L., Kleinberg, J., Lee, L., Danescu-Niculescu-Mizil, C., 2013. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In: Proceedings of the sixth ACM international conference on Web search and data mining. ACM, pp. 13–22.
- Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C., 2003. A neural probabilistic language model. *Journal of machine learning research* 3 (Feb), 1137–1155.
- Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks* 5 (2), 157–166.
- Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R., Bengio, S., 2015. Generating sentences from a continuous space. arXiv preprint arXiv:1511.06349.
- Branigan, H. P., Pickering, M. J., Cleland, A. A., 2000. Syntactic co-ordination in dialogue. *Cognition* 75 (2), B13–B25.
- Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42 (9), 2355–2368.
- Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R., 1994. Signature verification using a” siamese” time delay neural network. In: *Advances in Neural Information Processing Systems*. pp. 737–744.
- Chen, X., Qiu, X., Zhu, C., Liu, P., Huang, X., 2015. Long short-term memory neural networks for chinese word segmentation. In: *EMNLP*. pp. 1197–1206.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

- Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
- Church, K. W., 2000. Empirical estimates of adaptation: the chance of two noriegas is closer to $p/2$ than $p/2$. In: Proceedings of the 18th conference on Computational linguistics-Volume 1. Association for Computational Linguistics, pp. 180–186.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P., 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research* 12 (Aug), 2493–2537.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine learning* 20 (3), 273–297.
- Costa, A., Pickering, M. J., Sorace, A., 2008. Alignment in second language dialogue. *Language and Cognitive Processes* 23 (4), 528–556.
- Cui, Y., Chen, Z., Wei, S., Wang, S., Liu, T., Hu, G., 2016. Attention-over-attention neural networks for reading comprehension. arXiv preprint arXiv:1607.04423.
- Dai, A. M., Le, Q. V., 2015. Semi-supervised sequence learning. In: *Advances in Neural Information Processing Systems*. pp. 3079–3087.
- Danescu-Niculescu-Mizil, C., 2012. A computational approach to linguistic coordination. Ph.D. thesis, Cornell University.
- Danescu-Niculescu-Mizil, C., Lee, L., 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In: *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics, pp. 76–87.
- Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., Potts, C., 2013. No country for old members: User lifecycle and linguistic change in online communities. In: *Proceedings of the 22nd international conference on World Wide Web*. ACM, pp. 307–318.
- Dauphin, Y. N., Fan, A., Auli, M., Grangier, D., 2016. Language modeling with gated convolutional networks. arXiv preprint arXiv:1612.08083.

- Demonte, V., Soriano, O. F., 2005. Features in comp and syntactic variation: the case of '(de) queismo' in spanish. *Lingua* 115 (8), 1063–1082.
- Dos Santos, C. N., Gatti, M., 2014. Deep convolutional neural networks for sentiment analysis of short texts. In: *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*. pp. 69–78.
- Doyle, G., Frank, M. C., 2016. Investigating the sources of linguistic alignment in conversation. In: *Proceedings of the 54th Annual Conference of the Association for Computational Linguistics. (Volume 1: Long Papers)*. Vol. 1. pp. 526–536.
- Doyle, G., Goldberg, A., Srivastava, S., Frank, M., 2017. Alignment at work: Using language to distinguish the internalization and self-regulation components of cultural fit in organizations. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. pp. 603–612.
- Doyle, G., Yurovsky, D., Frank, M. C., 2016. A robust framework for estimating linguistic alignment in twitter conversations. In: *Proceedings of the 25th international conference on world wide web. International World Wide Web Conferences Steering Committee*, pp. 637–648.
- Elman, J. L., 1990. Finding structure in time. *Cognitive science* 14 (2), 179–211.
- Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., Tylén, K., 2012. Coming to terms: quantifying the benefits of linguistic coordination. *Psychological science* 23 (8), 931–939.
- Garrod, S., Pickering, M. J., 2009. Joint action, interactive alignment, and dialog. *Topics in Cognitive Science* 1 (2), 292–304.
- Gers, F. A., Schmidhuber, J., Cummins, F., 1999. Learning to forget: Continual prediction with lstm.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press, <http://www.deeplearningbook.org>.
- Graves, A., Jaitly, N., 2014. Towards end-to-end speech recognition with recurrent neural networks. In: *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*. pp. 1764–1772.

- Graves, A., Jaitly, N., Mohamed, A.-r., 2013a. Hybrid speech recognition with deep bidirectional lstm. In: Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on. IEEE, pp. 273–278.
- Graves, A., Mohamed, A.-r., Hinton, G., 2013b. Speech recognition with deep recurrent neural networks. In: Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on. IEEE, pp. 6645–6649.
- Graves, A., et al., 2012. Supervised sequence labelling with recurrent neural networks. Vol. 385. Springer.
- Gries, S. T., 2005. Syntactic priming: A corpus-based approach. *Journal of psycholinguistic research* 34 (4), 365–399.
- Gulcehre, C., Ahn, S., Nallapati, R., Zhou, B., Bengio, Y., 2016. Pointing the unknown words. *arXiv preprint arXiv:1603.08148*.
- Hadsell, R., Chopra, S., LeCun, Y., 2006. Dimensionality reduction by learning an invariant mapping. In: Computer vision and pattern recognition, 2006 IEEE computer society conference on. Vol. 2. IEEE, pp. 1735–1742.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural computation* 9 (8), 1735–1780.
- Jean, S., Cho, K., Memisevic, R., Bengio, Y., 2014. On using very large target vocabulary for neural machine translation. *arXiv preprint arXiv:1412.2007*.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia. ACM, pp. 675–678.
- Kalchbrenner, N., Grefenstette, E., Blunsom, P., 2014. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.
- Karpathy, A., Fei-Fei, L., 2015. Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3128–3137.

- Kaschak, M. P., Kutta, T. J., Jones, J. L., 2011. Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic bulletin & review* 18 (6), 1133–1139.
- Kim, Y., 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.
- Kim, Y., Jernite, Y., Sontag, D., Rush, A. M., 2016. Character-aware neural language models. In: *AAAI*. pp. 2741–2749.
- Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection.
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp. 1097–1105.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86 (11), 2278–2324.
- Li, H., 2017. Deep learning for natural language processing: advantages and challenges. *National Science Review*.
- Li, J., Ju, W., Nass, C., 2015a. Observer perception of dominance and mirroring behavior in human-robot relationships. In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, pp. 133–140.
- Li, J., Luong, M.-T., Jurafsky, D., 2015b. A hierarchical neural autoencoder for paragraphs and documents. arXiv preprint arXiv:1506.01057.
- Liu, P., Qiu, X., Huang, X., 2016. Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101.
- Liu, S., Yang, N., Li, M., Zhou, M., 2014. A recursive recurrent neural network for statistical machine translation.

- Lu, J., Yang, J., Batra, D., Parikh, D., 2016. Hierarchical question-image co-attention for visual question answering. In: Advances In Neural Information Processing Systems. pp. 289–297.
- Luong, M.-T., Pham, H., Manning, C. D., 2015. Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.
- Macías, C., Kim, H.-K., Yim, H.-J., Lee, K., 2013. 중급 스페인어 Español intermedio. SNU Press.
- Merity, S., Xiong, C., Bradbury, J., Socher, R., 2016. Pointer sentinel mixture models. arXiv preprint arXiv:1609.07843.
- Mikolov, T., 2010. Recurrent neural network based language model.
- Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013a. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J., 2013b. Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119.
- Mikolov, T., Yih, W.-t., Zweig, G., 2013c. Linguistic regularities in continuous space word representations.
- Mueller, J., Thyagarajan, A., 2016. Siamese recurrent architectures for learning sentence similarity.
- Nass, C., Lee, K. M., 2000. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '00. ACM, New York, NY, USA, pp. 329–336.
URL <http://doi.acm.org/10.1145/332040.332452>
- Navarro Colorado, B., Palomar, M., Martínez-Barco, P., et al., 2004. Automatic extraction of syntactic semantic patterns for multilingual resources.
- Noh, E., 2017a. Linguistic alignment in quasi-spoken corpus in spanish and marker list extraction modeling. In: Proceedings of the 2017 ACML workshop on Machine Learning for AI Platforms (MLAIP).

- Noh, E., 2017b. The structural and cognitive characteristics of linguistic alignment in the spanish spoken corpus. In: Proceedings of the 8th Annual Conference of Association of History, Literature, Science and Technology.
- Peng, H., Cambria, E., Zou, X., 2017. Radical-based hierarchical embeddings for chinese sentiment analysis at sentence level. In: Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference (FLAIRS).
- Pennebaker, J., Booth, R., Boyd, R., Francis, M., 2015. Linguistic inquiry and word count: Liwc2015.
- Pennington, J., Socher, R., Manning, C., 2014. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543.
- Pickering, M. J., Ferreira, V. S., 2008. Structural priming: A critical review. *Psychological bulletin* 134 (3), 427.
- Pickering, M. J., Garrod, S., 2004a. The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences* 27 (2), 212–225.
- Pickering, M. J., Garrod, S., 2004b. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27 (2), 169–190.
- Poria, S., Cambria, E., Hazarika, D., Majumder, N., Zadeh, A., Morency, L.-P., 2017. Context-dependent sentiment analysis in user-generated videos. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1. pp. 873–883.
- Prechelt, L., 2012. Early stopping—but when? In: *Neural networks: tricks of the trade*. Springer, pp. 53–67.
- Rabiner, L. R., 1989. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77 (2), 257–286.
- Raina, R., Madhavan, A., Ng, A. Y., 2009. Large-scale deep unsupervised learning using graphics processors. In: Proceedings of the 26th annual international conference on machine learning. ACM, pp. 873–880.

- Reitter, D., Keller, F., Moore, J. D., 2011. A computational cognitive model of syntactic priming. *Cognitive science* 35 (4), 587–637.
- Reitter, D., Moore, J. D., 2006. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In: *Proceedings of the Cognitive Science Society*. Vol. 28.
- Sak, H., Senior, A., Beaufays, F., 2014. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv preprint arXiv:1402.1128*.
- Santos, C. D., Zadrozny, B., 2014. Learning character-level representations for part-of-speech tagging. In: *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*. pp. 1818–1826.
- Santos, C. N. d., Guimaraes, V., 2015. Boosting named entity recognition with neural character embeddings. *arXiv preprint arXiv:1505.05008*.
- Schmidhuber, J., 2008. Learning complex, extended sequences using the principle of history compression. *Learning* 4 (2).
- Schuster, M., Paliwal, K. K., 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45 (11), 2673–2681.
- Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. Cnn features off-the-shelf: an astounding baseline for recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. pp. 806–813.
- Shen, Y., He, X., Gao, J., Deng, L., Mesnil, G., 2014. A latent semantic model with convolutional-pooling structure for information retrieval. In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, pp. 101–110.
- Shin, J.-Y., Cho, E. Y., Kim, U.-K., Macías, C., 2008. 스페인어 회화작문 Curso de Conversación y Redacción en Español. SNU Press.
- Shin, J.-Y., Lee, M.-K., Kim, U.-K., Blancafort, R., 2013a. 초급 스페인어 1 Español básico 1. SNU Press.

- Shin, J.-Y., Lee, M.-K., Kim, U.-K., Blancafort, R., 2013b. 초급 스페인어 2 Español básico 2. SNU Press.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 conference on empirical methods in natural language processing. pp. 1631–1642.
- Stenchikova, S., Stent, A., 2007. Measuring adaptation between dialogs. In: Proc. of the 8th SIGdial Workshop on Discourse and Dialogue.
- Sutskever, I., 2013. Training recurrent neural networks. University of Toronto, Toronto, Ont., Canada.
- Sutskever, I., Martens, J., Hinton, G. E., 2011. Generating text with recurrent neural networks. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11). pp. 1017–1024.
- Sutskever, I., Vinyals, O., Le, Q. V., 2014. Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp. 3104–3112.
- Tang, D., Qin, B., Liu, T., 2015. Document modeling with gated recurrent neural network for sentiment classification.
- Tong, E., Zadeh, A., Jones, C., Morency, L.-P., 2017. Combating human trafficking with deep multimodal models. arXiv preprint arXiv:1705.02735.
- Vinyals, O., Kaiser, Ł., Koo, T., Petrov, S., Sutskever, I., Hinton, G., 2015a. Grammar as a foreign language. In: Advances in Neural Information Processing Systems. pp. 2773–2781.
- Vinyals, O., Toshev, A., Bengio, S., Erhan, D., 2015b. Show and tell: A neural image caption generator. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3156–3164.
- Wang, Y., Reitter, D., Yen, J., 2014. Linguistic adaptation in conversation threads: Analyzing alignment in online health communities. In: Proceedings of the 2014 ACL Workshop on Cognitive Modeling and Computational Linguistics. pp. 55–62.

- Xie, Z., Wang, S. I., Li, J., Lévy, D., Nie, A., Jurafsky, D., Ng, A. Y., 2017. Data noising as smoothing in neural network language models. arXiv preprint arXiv:1703.02573.
- Xiong, C., Zhong, V., Socher, R., 2016. Dynamic coattention networks for question answering. arXiv preprint arXiv:1611.01604.
- Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y., 2015. Show, attend and tell: Neural image caption generation with visual attention. In: International Conference on Machine Learning. pp. 2048–2057.
- Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. J., Hovy, E. H., 2016. Hierarchical attention networks for document classification.
- Yih, S. W.-t., He, X., Meek, C., 2014. Semantic parsing for single-relation question answering.
- Yin, W., Kann, K., Yu, M., Schütze, H., 2017. Comparative study of cnn and rnn for natural language processing. arXiv preprint arXiv:1702.01923.
- Yin, W., Schütze, H., Xiang, B., Zhou, B., 2015. Abcnn: Attention-based convolutional neural network for modeling sentence pairs. arXiv preprint arXiv:1512.05193.
- Zadeh, A., Chen, M., Poria, S., Cambria, E., Morency, L.-P., 2017. Tensor fusion network for multimodal sentiment analysis. arXiv preprint arXiv:1707.07250.
- Zhang, Y., Wallace, B., 2015. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820.
- Zheng, X., Chen, H., Xu, T., 2013. Deep learning for chinese word segmentation and pos tagging. In: EMNLP. pp. 647–657.

Our github URL is <https://github.com/eunchung/Alignment>, containing twitter data and whole codes for this dissertation.

A.1 Crawling and Preprocessing - Whole codes

Code 1: Crawler.py

```
import os
import re
import tweepy
import time
from tweepy import OAuthHandler

class TwitterClient(object):
    """
    Generic Twitter Class for sentiment analysis.
    """
    def __init__(self):
        """
        Class constructor or initialization method.
        """
        # keys and tokens from the Twitter Dev Console
        consumer_key = '트위터에서 승인받은 정보를 입력'
        consumer_secret = '트위터에서 승인받은 정보를 입력'
        access_token = '트위터에서 승인받은 정보를 입력'
        access_token_secret = '트위터에서 승인받은 정보를 입력'

        # attempt authentication
        try:
            # create OAuthHandler object
            self.auth = OAuthHandler(consumer_key, consumer_secret)
            # set access token and secret
            self.auth.set_access_token(access_token,
                                      access_token_secret)
            # create tweepy API object to fetch tweets
            self.api = tweepy.API(self.auth)
        except:
            print("Error: Authentication Failed")

    def get_tweets(self, query, count = 10, RT_number = 0):
        """
        Main function to fetch tweets and parse them.
        """
        # empty list to store parsed tweets
        tweets = []

        try:
            filtered_fetched_tweets = []
            filtered_fetched_tweet_ids = []
            # call twitter api to fetch tweets. count 만큼 lang="es"
            # 서치하고, 그 중에서 리트윗횟수 높은거만 추출하자.
            tweets = self.api.search(query, count=count, lang="es",
                                     filter="retweets")
```

```

        fetched_tweets = self.api.search(q = query, count = count,
                                         lang="es")
        # 그냥 id_str 이 아니라, retweeted_status.id_str 을 써야,
        # 제대로 된 트위터 status 아이디로 감.

        # 리트윗 수 높은 것만 추출 .
        for tweet in fetched_tweets:
            if tweet.retweet_count >= RT_number and tweet.text:
                #not in filtered_fetched_tweets:
                filtered_fetched_tweets.append(tweet.text)
                filtered_fetched_tweet_ids.append(tweet.retweeted_status.id_str)

        # parsing tweets one by one
        for tweet, id in zip(filtered_fetched_tweets,
                             filtered_fetched_tweet_ids) :
            # empty dictionary to store required params of a tweet
            parsed_tweet = {}

            # saving text of tweet
            parsed_tweet['raw_text'] = tweet.strip()
            parsed_tweet['tweet_id'] = id
            tweets.append(parsed_tweet)

        # return parsed tweets
        return tweets

    except tweepy.TweepError as e:
        print("Error : " + str(e))

def get_replies(self, query, tweet_id, count):
    """
    트위터의 reply 들을 얻기위한 코드
    참조 :
        https://stackoverflow.com/questions/29928638/getting-tweet-replies-to-a-par

    그냥 tweet.id_str 이 아니라, tweet.retweeted_status.id_str 을
    써야, 제대로 된 트위터 status 아이디로 감. !!
    """
    replies = []

    try:
        filtered_fetched_replies = []
        fetched_tweets = self.api.search(q = query, since_id =
                                         tweet_id, count = count, lang="es")

        # 그냥 리트윗은 제외하고 , 리트윗에 맨션 있는것만 추출 .
        for tweet in fetched_tweets:
            if tweet.in_reply_to_status_id==int(tweet_id):
                filtered_fetched_replies.append(tweet.text)

        # parsing tweets one by one
        for reply in filtered_fetched_replies:
            parsed_reply = {}
            parsed_reply['raw_text'] = reply.strip()
            replies.append(parsed_reply)
    
```

```

        # return parsed tweets
        return replies

    except tweepy.TweepError as e:
        print("Error : " + str(e))

def main():
    # creating object of TwitterClient Class
    api = TwitterClient()
    count = 100
    query = 'RT' #independencia'
    RT_number = 100

    if not os.path.exists('./data/tweet/crawled'):
        os.makedirs('./data/tweet/crawled')

    print(query)

    count_file = 0
    for _ in range(100):
        count_file +=1
        if count_file % 5 == 0:
            print('crawled %s times' % count_file)

        time_now = time.time()
        with open('./data/tweet/crawled/' + query + '_' +
            str(RT_number)+'_'+str(time_now)+'.txt',
            'w',encoding='UTF-8') as w:
            # calling function to get tweets
            w.write('RT 로 리트윗 검색. 리트윗 100개 넘는 것들 + 그것에
                달린 reply 들 기록. \n')
            tweets = api.get_tweets(query, count, RT_number)

            for tweet in tweets:
                query_reply = tweet['raw_text'].split()[1][:-1]
                replies = api.get_replies(query_reply,
                    tweet['tweet_id'],100)

                if replies:
                    w.write(tweet['raw_text']+'\n')
                    for reply in replies:
                        w.write(reply['raw_text']+'\n')
                    w.write('\n')

            # wait 3*60 sec to avoid 'Rate limit exceeded'
            time.sleep(3*60)

if __name__ == "__main__":
    # calling main function
    main()

```

Code 2: arrange_retweet.py

```

import sys
import os

def main():
    output = './data/tweet/total_retweet.txt'

    count = 0 # tweet/retweet pair 수 세기
    with open(output, 'w', encoding='utf-8') as w:

        for root, dirs, files in os.walk("./data/tweet/crawled/",
            topdown = False):
            for name in files:
                print(os.path.join(root, name))
                with open(os.path.join(root, name), 'r', encoding =
                    'utf-8') as r:
                    sentence_continue = ""
                    finished_list = ['empty']
                    next_is_Blank = False
                    for sentence in r.readlines():
                        count += 1
                        if count % 100 == 0:
                            print(count)

                        if sentence.strip() == "":
                            next_is_Blank = True
                            count -= 1 # 빈칸은 세지않음 .

                        # tweet 사이에 개행이 있는 tweet 는 sentence_continue
                        # 로 두고, 개행들 다 합침.
                        if sentence.strip() != "" and '@' not in sentence:
                            sentence_continue = sentence_continue +
                                sentence.strip()
                            count -= 1 # 개행되어 있는 건 하나의 트윗으로
                                # 세기때문에 count 하지않음 .
                            continue
                        elif sentence.strip() != "":
                            sentence_finished = sentence_continue
                            sentence_continue = sentence.strip()
                            finished_list.append(sentence_finished)

                    if next_is_Blank:
                        next_is_Blank = False
                        finished_list.append('empty')
                        count -= 1 # empty 표시는 세지않음.

    next_is_RT = False
    finished_list.pop(1) # 한글 한 줄 빼기
    count -= 1
    for sentence in finished_list:
        if sentence == 'empty':
            next_is_RT = True
            continue
        if next_is_RT == True:
            RT = sentence
            next_is_RT = False
            continue

```

```

        w.write(RT+'\t'+sentence+'\n')
        count -= 1 # 파일의 마지막 빈 한글 지움.
    print('Total number of tweet-retweet pair:',count) #미세하게
    다룰때가 있네..

if __name__ == "__main__":
    main()

```

As we mentioned at section 3.2.2, we removed account(accountname), hashtag (#something), 'RT', links, special characters to clean tweet text. The python library 'Regular Expression' is utilized for preprocessing.

Code 3: Preprocessing code with library *Regular Expression (RE)*

```

def clean_tweet(tweet):
    return ' '.join(re.sub("([A-Za-z0-9_]+)|#[A-Za-z0-9]+|(RT)|(\w+:\w+\w\S+)|([\^A-Za-z0-9_áÁéÉííóóúÚñÑüÜ°°€; !?¿(){}<>,...@#%& '\\"

```

We used ToktokTokenizer, a tokenizer for Spanish. When using ToktokTokenizer only, the 'word' + '.' sometimes are not be separated properly. For example, 'done.' should be separated into 'done' + '.' but they were not separated. Thus we additionally used nltk.wordpunct_tokenize.¹

Code 4: Tokenizing with library *Natural Language Tool Kit (NLTK)*

```

import nltk
from nltk.tokenize.toktok import ToktokTokenizer
toktok = ToktokTokenizer()
word_list = toktok.tokenize('
    '.join(nltk.wordpunct_tokenize(sentence)))

```

Then, we separate tweet corpus into train, valid and test data. 80% of the corpus belongs to train data, 10% of the corpus belongs to valid data and 10% of the corpus belongs to test data².

Code 5: Separating train data into a training set and a validation set with library

¹Reference site: <https://github.com/nltk/nltk/issues/1558>

²Reference site: http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

scikit-learn

```
from sklearn.model_selection import train_test_split
corpus = data.readlines()
train, valid = train_test_split(corpus, test_size=0.1,
                                random_state=6)
```

Code 6: split_train_valid.py

```
import sys
from sklearn.model_selection import train_test_split

output_train = 'train_' + sys.argv[1]
output_valid = 'valid_' + sys.argv[1]
#output_test = 'test_twitter.txt'
#output_test = 'test_text.txt'
#output_test = 'test_all.txt' # 이걸 위의 두 데이터를 합쳐서 만듬.

import re
import nltk
from nltk.tokenize.toktok import ToktokTokenizer
toktok = ToktokTokenizer()

import unicodedata
import string

# 윈도우에서 스페인어 보기 위해 .
# Turn a Unicode string to plain ASCII, thanks to
# http://stackoverflow.com/a/518232/2809427
def unicodeToAscii(s):
    all_letters = string.ascii_letters + " .,;'"
    return ''.join(
        c for c in unicodedata.normalize('NFD', s)
        if unicodedata.category(c) != 'Mn'
        and c in all_letters
    )

def clean_tweet(tweet):
    """
    Utility function to clean tweet text by removing links, special
    characters
    using simple regex statements.
    """
    return '
'.join(re.sub("(@[A-Za-z0-9_]+)|(#[A-Za-z0-9_]+)|(RT)|(\w+:\/\/\/S+)|([^\A-Za-z0-9_!@&quot;'\n\.\,\;\-\_\+\*\'\\"^])", " ", tweet).split()) # 최종버전

with open(output_train, 'w', encoding='utf-8') as w_train, \
    open(output_valid, 'w', encoding='utf-8') as w_valid, \
    open(output_test, 'w', encoding='utf-8') as w_test, \
    open('./'+sys.argv[1], 'r', encoding='utf-8') as data, \
    open('./'+sys.argv[2], 'r', encoding='utf-8') as test_data:

    corpus = data.readlines()
    clean_tokenized_corpus = []
```



```

for i, line in enumerate(corpus):
    #print(unicodeToAscii(line))
    sent1, sent2, label = line.strip().split('\t')
    sent1 = clean_tweet(sent1)
    sent2 = clean_tweet(sent2)
    if (len(sent1) == 0 or len(sent2) == 0):
        print('[train_data] %d-th sentence is empty after cleaning'
              % (i+1))
        continue
    word_list_1 = toktok.tokenize('
        '.join(nltk.wordpunct_tokenize(sent1)))
    word_list_2 = toktok.tokenize('
        '.join(nltk.wordpunct_tokenize(sent2)))
    sent1=' '.join(word_list_1)
    sent2=' '.join(word_list_2)
    clean_tokenized_corpus.append(sent1.strip()+'\t'+sent2.strip()+'\t'+
                                  label.strip() +'\n')

train, valid = train_test_split(clean_tokenized_corpus,
                                test_size=0.1, random_state=6)

print('length of train_data', len(train))
print('length of valid_data', len(valid))

for train_sent in train:
    sent1, sent2, label = train_sent.strip().split('\t')
    w_train.write(sent1.strip()+' endofsentence\t'+sent2.strip()+'
        endofsentence\t'+ label.strip() +'\n')
    #w_train.write(sent1.strip()+'\t'+sent2.strip()+'\t'+
        label.strip() +'\n') # 나중에 augmented 할때 endofsentence
        추가하는 방식을 사용. 이때 추가해두면 worddropout 시에 unk 되버림.

for valid_sent in valid:
    sent1, sent2, label = valid_sent.strip().split('\t')
    w_valid.write(sent1.strip()+' endofsentence\t'+sent2.strip()+'
        endofsentence\t'+ label.strip() +'\n')

test_corpus = test_data.readlines()
for i, line in enumerate(test_corpus):
    sent1, sent2, label = line.strip().split('\t')
    sent1 = clean_tweet(sent1)
    sent2 = clean_tweet(sent2)
    if (len(sent1) == 0 or len(sent2) == 0):
        print('[test_data] %d-th sentence is empty after cleaning' %
              (i+1))
        continue
    word_list_1 = toktok.tokenize('
        '.join(nltk.wordpunct_tokenize(sent1)))
    word_list_2 = toktok.tokenize('
        '.join(nltk.wordpunct_tokenize(sent2)))
    sent1=' '.join(word_list_1)
    sent2=' '.join(word_list_2)
    w_test.write(sent1.strip()+' endofsentence\t'+sent2.strip()+'
        endofsentence\t'+ label.strip() +'\n')

```

Code 7: arrange_data_no_label.py

```
import sys
import re
import nltk
from nltk.tokenize.toktok import ToktokTokenizer
toktok = ToktokTokenizer()

import unicodedata
import string

# 윈도우에서 스페인어 보기 위해 .
# Turn a Unicode string to plain ASCII, thanks to
# http://stackoverflow.com/a/518232/2809427
def unicodeToAscii(s):
    all_letters = string.ascii_letters + " .,;'"
    return ''.join(
        c for c in unicodedata.normalize('NFD', s)
        if unicodedata.category(c) != 'Mn'
        and c in all_letters
    )

def clean_tweet(tweet):
    """
    Utility function to clean tweet text by removing links, special
    characters
    using simple regex statements.
    """
    return '
'.join(re.sub("(@[A-Za-z0-9_]+)|(#[A-Za-z0-9_]+)|(RT)|(\w+:\/\/\/S+)|([^\A-Za-z0-9_!@&quot;'\`~\:\;\-\_\+\*\'\`\\"^])", " ", tweet).split()) # 최종버전

def main():
    output = 'preprocessed_' + sys.argv[1]
    count = 0
    with open(output, 'w', encoding='utf-8') as w,
        open('./'+sys.argv[1], 'r', encoding='utf-8') as tweet:

        tweet_corpus = tweet.readlines()
        for i, line in enumerate(tweet_corpus):
            sent1, sent2 = line.strip().split('\t')
            sent1 = clean_tweet(sent1)
            sent2 = clean_tweet(sent2)
            if (len(sent1) == 0 or len(sent2) == 0):
                print('[twitter_crawled_data] %d-th sentence is empty
                    after cleaning' % (i+1))
                continue
            word_list_1 = toktok.tokenize('
'.join(nltk.wordpunct_tokenize(sent1)))
            word_list_2 = toktok.tokenize('
'.join(nltk.wordpunct_tokenize(sent2)))
            sent1=' '.join(word_list_1)
            sent2=' '.join(word_list_2)
            w.write(sent1.strip()+' endofsentence\t'+sent2.strip()+'
                    endofsentence\ttemp_label\n')

        count += 1
```

```

        if count % 100 == 0:
            print(count)

if __name__ == "__main__":
    main()

```

A.2 Alignment Classifier - Whole codes

Code 8: baseline.py

```

import sys
import unicodedata
import string
import os

test_data_path = sys.argv[1]
directory_name = sys.argv[2]
alignment_marker_type = sys.argv[3]

# 윈도우에서 스페인어 보기 위해 .
# Turn a Unicode string to plain ASCII, thanks to
# http://stackoverflow.com/a/518232/2809427
def unicodeToAscii(s):
    all_letters = string.ascii_letters + " .,;'"
    return ''.join(
        c for c in unicodedata.normalize('NFD', s)
        if unicodedata.category(c) != 'Mn'
        and c in all_letters
    )

# 대표하는 alignment markers = 이 단어가 후속 문장에 포함되면
# alignment 다 라고 생각가능한 마커만 뽑음 .
def marker_set(x):
    return {
        '1': ['el', 'la', 'los', 'las', 'lo'],
        '2': ['el', 'la', 'los', 'las', 'lo', 'sí', 'bien', 'si',
              'mucho', 'no'],
        '3': ['el', 'la', 'los', 'las', 'lo', 'sí', 'bien', 'si',
              'mucho', 'no', 'con', 'a', 'en', 'por', 'hay'],
        '4': ['el', 'la', 'los', 'las', 'lo', 'sí', 'bien', 'si',
              'mucho', 'no', 'con', 'a', 'en', 'por', 'hay', 'siempre',
              'nunca', 'y', 'pero', 'así'],
        'd1': ['de', 'que', 'a', 'la', 'y'],
        'd2': ['de', 'que', 'a', 'la', 'y', 'no', 'el', 'en', 'es',
              'los'],
        'd3': ['de', 'que', 'a', 'la', 'y', 'no', 'el', 'en', 'es',
              'los', 'por', 'lo', 'un', 'se', 'con'],
        'd4': ['de', 'que', 'a', 'la', 'y', 'no', 'el', 'en', 'es',
              'los', 'por', 'lo', 'un', 'se', 'con', 'me', 'si', 'te',
              'para', 'del'],
        'd5': ['de', 'que', 'a', 'la', 'y', 'no', 'el', 'en', 'es',
              'los', 'por', 'lo', 'un', 'se', 'con', 'me', 'si', 'te',

```

```

        'para', 'del', \
        'una', 'las', 'mi', 'pero', 'al', 'más', 'le', 'como', 'ya', 'yo', 'q', 'todo', 'gra
        'nos', 'tu', 'o', 'esta', 'hoy', 'muy', 'ha', 'eso', 'ni', 'porque'],
        'differ': ['yo', 'todo', 'gracias', 'esto', 'nos', 'muy', 'eso', 'ni'],
    }.get(x)

alignment_markers = marker_set(alignment_marker_type)
print([unicodeToAscii(marker) for marker in alignment_markers])

# Test baseline
correct = 0
total = 0
if not os.path.exists('./result'):
    os.makedirs('./result')
if not os.path.exists('./result/'+directory_name):
    os.makedirs('./result/'+directory_name)

with
    open('./result/'+directory_name+'/baseline_maker_%s.txt'%alignment_marker_type,
        'w', encoding='utf-8') as w, open(test_data_path, 'r',
        encoding='utf-8') as test_data:

    for line in test_data.readlines():
        sentence_1, sentence_2, label =
            line.lower().strip().split('\t')
        word_list = sentence_2.split() # 후속 문장의 단어만 보고 판단.

        predicted = 2 # 기본은 none (2) 이고, marker 가 포함되었을때만
            alignment (1) 로 함.
        for candidate in alignment_markers:
            if candidate in word_list:
                predicted = 1
                break

        total += 1

        if int(predicted) != int(label):
            w.write('label\t' + str(label) + '\t' + 'predicted\t' +
                str(predicted) + '\t' + sentence_1 + '\t' + sentence_2
                + '\n')
            correct += (int(predicted) == int(label))

    print('Test Accuracy of baseline: %0.2f %%' % (100 * correct /
        total))
    w.write('[Test Accuracy of baseline] : %0.2f %% \n\n' % (100 *
        correct / total))

os.rename('./result/'+directory_name+'/baseline_maker_%s.txt'%alignment_marker_type,
    './result/'+directory_name+'/baseline_maker_%s_Acc%0.2f.txt'%(alignment_marker_t
    (100 * correct / total)))

```

Code 9: augment_word_dropout_separately.py

```
import sys
```

```

import numpy as np

import unicodedata
import string

# 윈도우에서 스페인어 보기 위해 .
# Turn a Unicode string to plain ASCII, thanks to
# http://stackoverflow.com/a/518232/2809427
def unicodeToAscii(s):
    all_letters = string.ascii_letters + " .,;'"
    return ''.join(
        c for c in unicodedata.normalize('NFD', s)
        if unicodedata.category(c) != 'Mn'
        and c in all_letters
    )

word_dropout_rate = 0.1 # 0.2

#output = 'WD_%0.2f_aug_train_twitter.txt'%word_dropout_rate
output = 'WD_%0.2f_aug_train_text.txt'%word_dropout_rate

number_of_augmented_align = 2
number_of_augmented_none = 8

with open(output, 'w', encoding='utf-8') as w_augmented, \
    open('./'+sys.argv[1], 'r', encoding='utf-8') as align, \
    open('./'+sys.argv[2], 'r', encoding='utf-8') as none:
    align_corpus = align.readlines()
    none_corpus = none.readlines()

    # 본래 align_data
    for align_sent in align_corpus:
        w_augmented.write(align_sent)

    # number_of_augmented_align - 1 만큼 augmented 한 데이터를 추가함.
    for _ in range(number_of_augmented_align - 1):
        for align_sent in align_corpus:
            #print(unicodeToAscii(train_sent))
            sent1, sent2, label = align_sent.strip().split('\t')

            word_list_1 = sent1.split()
            replace_index_1 = np.random.rand(len(word_list_1)) <
                word_dropout_rate # will return list of ture / false. ex)
                [True False True]
            replaced_word_list_1 = []
            for word, replace in zip(word_list_1, replace_index_1):
                if replace:
                    word = 'unk'
            replaced_word_list_1.append(word)
            sent1=' '.join(replaced_word_list_1)
            #print(unicodeToAscii(sent1))

            word_list_2 = sent2.split()
            replace_index_2 = np.random.rand(len(word_list_2)) <
                word_dropout_rate # will return list of ture / false. ex)
                [True False True]
            replaced_word_list_2 = []

```

```

for word, replace in zip(word_list_2, replace_index_2):
    if replace:
        word = 'unk'
        replaced_word_list_2.append(word)
    sent2=' '.join(replaced_word_list_2)
    #print(unicodeToAscii(sent2))

w_augmented.write(sent1.strip()+'\t'+sent2.strip()+'\t'+
    label.strip() +'\n')

# 본래 train_data
for none_sent in none_corpus:
    w_augmented.write(none_sent)

# number_of_augmented_none - 1 만큼 augmented 한 데이터를 추가함.
for _ in range(number_of_augmented_none - 1):
    for none_sent in none_corpus:
        #print(unicodeToAscii(train_sent))
        sent1, sent2, label = none_sent.strip().split('\t')

        word_list_1 = sent1.split()
        replace_index_1 = np.random.rand(len(word_list_1)) <
            word_dropout_rate # will return list of ture / false. ex)
            [True False True]
        replaced_word_list_1 = []
        for word, replace in zip(word_list_1, replace_index_1):
            if replace:
                word = 'unk'
                replaced_word_list_1.append(word)
        sent1=' '.join(replaced_word_list_1)
        #print(unicodeToAscii(sent1))

        word_list_2 = sent2.split()
        replace_index_2 = np.random.rand(len(word_list_2)) <
            word_dropout_rate # will return list of ture / false. ex)
            [True False True]
        replaced_word_list_2 = []
        for word, replace in zip(word_list_2, replace_index_2):
            if replace:
                word = 'unk'
                replaced_word_list_2.append(word)
        sent2=' '.join(replaced_word_list_2)
        #print(unicodeToAscii(sent2))

        w_augmented.write(sent1.strip()+'\t'+sent2.strip()+'\t'+
            label.strip() +'\n')

```

Code 10: main.py

```

import torch
import torch.nn as nn
import torch.nn.functional as F

from torch.utils.data import Dataset, DataLoader

```

```

from torch.autograd import Variable

import sys
import numpy
import os

from utils import *
from models import *

# Hyper Parameters
max_sequence_length = 30
max_vocabulary_size = 25000
embedding_size = 300
hidden_size = int(sys.argv[6])#256
num_layers = 2
num_classes = 2
batch_size = 150
num_epochs = 30
learning_rate = 0.001
dropout_rate = float(sys.argv[7]) #0

train_data_path = sys.argv[1]
valid_data_path = sys.argv[2]
test_data_path = sys.argv[3]
directory_name = sys.argv[4] #'Twit_CNN'
model_name = sys.argv[5] #'CNN'

word_to_ix, ix_to_word, vocab_size =
    make_or_load_dict(train_data_path, character=False)

def model(x):
    return {
        'BiLSTM': BiLSTM(vocab_size, embedding_size, hidden_size,
            num_layers, num_classes, dropout_rate),
        'CNN': CNN(vocab_size, embedding_size, num_classes,
            dropout_rate, kernel_num=hidden_size),
        'Cha_CNN_LSTM': Cha_CNN_LSTM(vocab_size, embedding_size,
            num_classes, dropout_rate, kernel_num =hidden_size),
        'Siamese_BiLSTM': Siamese_BiLSTM(vocab_size, embedding_size,
            hidden_size, num_layers, num_classes, dropout_rate),
        'Siamese_CNN': Siamese_CNN(vocab_size, num_classes,
            embedding_size, kernel_num = hidden_size),
    }.get(x)

model = model(model_name)
model.cuda()
print(model)

if model_name in ['BiLSTM', 'CNN', 'Cha_CNN_LSTM'] :
    train_dataset = AlignmentDataset(train_data_path, word_to_ix,
        batch_size)
    valid_dataset = AlignmentDataset(valid_data_path, word_to_ix, 1)
    test_dataset = AlignmentDataset(test_data_path, word_to_ix, 1)
elif model_name in ['Siamese_BiLSTM', 'Siamese_CNN']:
    train_dataset = AlignmentDataset_seperate_sent(train_data_path,
        word_to_ix, batch_size)
    valid_dataset = AlignmentDataset_seperate_sent(valid_data_path,

```

```

        word_to_ix, 1)
    test_dataset = AlignmentDataset_seperate_sent(test_data_path,
        word_to_ix, 1)

train_loader = DataLoader(dataset=train_dataset, shuffle=True)
valid_loader = DataLoader(dataset=valid_dataset)
test_loader = DataLoader(dataset=test_dataset)
print('length of train dataset', len(train_dataset)*batch_size)
print('length of valid dataset', len(valid_dataset))
print('length of test dataset', len(test_dataset))

# Loss and Optimizer
loss_function = nn.CrossEntropyLoss() # = log softmax + NLL loss
optimizer = torch.optim.Adam(model.parameters(), lr=learning_rate)

print_every = (len(train_dataset)/2)*2
losses = 0
keep_valid_loss = 1e06
keep_valid_accuracy = 0
if not os.path.exists('./models'):
    os.makedirs('./models')
if not os.path.exists('./models/'+directory_name):
    os.makedirs('./models/'+directory_name)

for epoch in range(num_epochs): # again, normally you would NOT do
    300 epochs, it is toy data
    if model_name in ['BiLSTM', 'CNN', 'Cha_CNN_LSTM']:
        for i, (sentences, labels) in enumerate(train_loader):
            model.zero_grad()
            logits = model(Variable(sentences.view(batch_size,
                -1)).cuda())
            loss = loss_function(logits,
                Variable(labels.view(batch_size)).cuda())
            loss.backward()
            torch.nn.utils.clip_grad_norm(model.parameters(), 1)
            optimizer.step()
            losses += loss.data[0]

        if ((i+1)+epoch*len(train_dataset)) % print_every == 0:
            # valid set test
            valid_loss = 0
            valid_correct = 0
            valid_total = 0
            for sentence, label in valid_loader:
                output = model(Variable(sentence.view(1,
                    -1)).cuda(), train=False)
                loss = loss_function(output,
                    Variable(label.view(1)).cuda())
                valid_loss += loss.data[0]

            _, predicted = torch.max(output.data, 1) # 두번째
            아웃풋 값은 argmax 를 반환
            valid_total += label.size(0)
            valid_correct += (predicted.cpu() == label).sum()
            valid_accuracy = (100 * valid_correct / valid_total)

```



```

# test set test
test_loss = 0
test_correct = 0
test_total = 0
for sentence, label in test_loader:
    output = model(Variable(sentence.view(1,
        -1)).cuda(), train=False)
    loss = loss_function(output,
        Variable(label.view(1)).cuda())
    test_loss += loss.data[0]

    _, predicted = torch.max(output.data, 1) # 두번째
    아웃풋 값은 argmax 를 반환
    test_total += label.size(0)
    test_correct += (predicted.cpu() == label).sum()
test_accuracy = (100 * test_correct / test_total)

if valid_accuracy > keep_valid_accuracy:
    keep_valid_accuracy = valid_accuracy
# if valid_loss < keep_valid_loss:
#     keep_valid_loss = valid_loss
best_valid_epoch = epoch + 1
torch.save(model.state_dict(),
    './models/'+directory_name+'/%s_hid%d_D%0.2f_epoch_%d_best_valid_%.4f'
    % (model_name, hidden_size, dropout_rate,
    (epoch+1)))
print ('Ep [%d/%d], Step [%d/%d], L: %.4f, V_L:
    %.4f, V_ACC: %0.2f %, Te_L: %.4f, Te_ACC: %0.2f
    %% <best valid>'
    %(epoch+1, num_epochs,
    (i+1)+epoch*len(train_dataset),
    num_epochs*len(train_dataset), losses,
    valid_loss, valid_accuracy, test_loss,
    test_accuracy))

else:
    torch.save(model.state_dict(),
        './models/'+directory_name+'/%s_hid%d_D%0.2f_epoch_%d.pkl'
        % (model_name, hidden_size, dropout_rate,
        (epoch+1)))
    print ('Ep [%d/%d], Step [%d/%d], L: %.4f, V_L:
        %.4f, V_ACC: %0.2f %, Te_L: %.4f, Te_ACC: %0.2f
        %%'
        %(epoch+1, num_epochs,
        (i+1)+epoch*len(train_dataset),
        num_epochs*len(train_dataset), losses,
        valid_loss, valid_accuracy, test_loss,
        test_accuracy))
losses = 0

elif model_name in ['Siamese_BiLSTM', 'Siamese_CNN']:
    for i, (sentence_1, sentence_2, labels) in
        enumerate(train_loader):
            model.zero_grad()
            logits = model(Variable(sentence_1.view(batch_size,
                -1)).cuda(), Variable(sentence_2.view(batch_size,

```

```

-1)).cuda())
loss = loss_function(logits,
    Variable(labels.view(batch_size)).cuda())
loss.backward()
torch.nn.utils.clip_grad_norm(model.parameters(), 1)
optimizer.step()
losses += loss.data[0]

if ((i+1)+epoch*len(train_dataset)) % print_every == 0:
    # valid set test
    valid_loss = 0
    valid_correct = 0
    valid_total = 0
    for sentence_1, sentence_2, label in valid_loader:
        output = model(Variable(sentence_1.view(1,
            -1)).cuda(), Variable(sentence_2.view(1,
            -1)).cuda(), train=False)
        loss = loss_function(output,
            Variable(label.view(1)).cuda())
        valid_loss += loss.data[0]

        _, predicted = torch.max(output.data, 1) # 두번째
        아웃풋 값은 argmax 를 반환
        valid_total += label.size(0)
        valid_correct += (predicted.cpu() == label).sum()
    valid_accuracy = (100 * valid_correct / valid_total)

    # test set test
    test_loss = 0
    test_correct = 0
    test_total = 0
    for sentence_1, sentence_2, label in test_loader:
        output = model(Variable(sentence_1.view(1,
            -1)).cuda(), Variable(sentence_2.view(1,
            -1)).cuda(), train=False)
        loss = loss_function(output,
            Variable(label.view(1)).cuda())
        test_loss += loss.data[0]

        _, predicted = torch.max(output.data, 1) # 두번째
        아웃풋 값은 argmax 를 반환
        test_total += label.size(0)
        test_correct += (predicted.cpu() == label).sum()
    test_accuracy = (100 * test_correct / test_total)

    if valid_accuracy > keep_valid_accuracy:
        keep_valid_accuracy = valid_accuracy
    # if valid_loss < keep_valid_loss:
    #     keep_valid_loss = valid_loss
    best_valid_epoch = epoch + 1
    torch.save(model.state_dict(),
        './models/'+directory_name+'/%s_hid%d_D%0.2f_epoch_%d_best_valid_a
        % (model_name, hidden_size, dropout_rate,
        (epoch+1)))
    print ('Ep [%d/%d], Step [%d/%d], L: %.4f, V_L:
        %.4f, V_ACC: %0.2f %, Te_L: %.4f, Te_ACC: %0.2f
        %% <best valid>')

```

```

        %(epoch+1, num_epochs,
          (i+1)+epoch*len(train_dataset),
          num_epochs*len(train_dataset), losses,
          valid_loss, valid_accuracy, test_loss,
          test_accuracy))

    else:
        torch.save(model.state_dict(),
          './models/'+directory_name+'/%s_hid%d_D%0.2f_epoch_%d.pkl'
          % (model_name, hidden_size, dropout_rate,
            (epoch+1)))
        print ('Ep [%d/%d], Step [%d/%d], L: %.4f, V_L:
          %.4f, V_ACC: %0.2f %, Te_L: %.4f, Te_ACC: %0.2f
          %')
        %(epoch+1, num_epochs,
          (i+1)+epoch*len(train_dataset),
          num_epochs*len(train_dataset), losses,
          valid_loss, valid_accuracy, test_loss,
          test_accuracy))
    losses = 0

# Test the Model
correct = 0
total = 0
missed_pairs = []
if model_name in ['BiLSTM', 'CNN', 'Cha_CNN_LSTM']:
    for sentence, label in test_loader:

        outputs = model(Variable(sentence.view(1, -1)).cuda(),
          train=False)
        _, predicted = torch.max(outputs.data, 1) # 두번째 아웃풋 값은
          argmax 를 반환

        total += label.size(0) # batch 쓰는 경우.
        if (predicted.cpu().numpy() != label.numpy()):
            sentence = [ix_to_word[word_idx] for word_idx in
              sentence.long()[0][0].numpy()]
            missed_pairs.append('label: ' + str(label.numpy()[0][0])
              + '\t' + 'predicted: ' + str(predicted.cpu().numpy()) +
              '\t' + ' '.join(sentence)+'\n')
        correct += (predicted.cpu() == label).sum()

elif model_name in ['Siamese_BiLSTM', 'Siamese_CNN']:
    for sentence_1, sentence_2, label in test_loader:

        output = model(Variable(sentence_1.view(1, -1)).cuda(),
          Variable(sentence_2.view(1, -1)).cuda(), train=False)
        _, predicted = torch.max(output.data, 1) # 두번째 아웃풋 값은
          argmax 를 반환

        total += label.size(0) # batch 쓰는 경우.
        if (predicted.cpu().numpy() != label.numpy()):
            sentence_1 = [ix_to_word[word_idx] for word_idx in
              sentence_1.long()[0][0].numpy()]
            sentence_2 = [ix_to_word[word_idx] for word_idx in
              sentence_2.long()[0][0].numpy()]

```

```

        missed_pairs.append('label: ' + str(label.numpy()[0][0])
                              + '\t' + 'predicted: ' + str(predicted.cpu().numpy()) +
                              '\t' + ' '.join(sentence_1) + '\t' +
                              ' '.join(sentence_2) + '\n')
        correct += (predicted.cpu() == label).sum()

print('Test Accuracy of the model: %0.2f %%' % (100 * correct /
        total))
# write result and save model
if not os.path.exists('./result'):
    os.makedirs('./result')
if not os.path.exists('./result/'+directory_name):
    os.makedirs('./result/'+directory_name)

with
    open('./result/'+directory_name+'/%s_hid%d_D%0.2f_Acc%0.2f.txt'
          % (model_name, hidden_size, dropout_rate, 100 * correct /
              total), 'a', encoding='utf-8') as w:

    w.write('[setting]:
            '+'\tbatch_size\t'+str(batch_size)+'\temb_size\t'+str(embedding_size)+'\tHid\t'
    w.write('[Test Accuracy of the model]: %0.2f %% \n' % (100 *
        correct / total))
    w.write('[saved to]: ./models/%s_hid%d_D%0.2f_Acc%0.2f.pkl\n' %
        (model_name, hidden_size, dropout_rate, 100 * correct /
            total))
    torch.save(model.state_dict(),
        './models/'+directory_name+'/%s_hid%d_D%0.2f_Acc%0.2f_last_epoch.pkl'
        % (model_name, hidden_size, dropout_rate, 100 * correct /
            total))

    for miss in missed_pairs:
        w.write(miss)
    w.write('--*100+\n\n')

# load the best valid model.
correct = 0
total = 0
missed_pairs = []
#best_valid_epoch = 10
model.load_state_dict(torch.load('./models/'+directory_name+'/%s_hid%d_D%0.2f_epoch%0.2f.pkl'
    % (model_name, hidden_size, dropout_rate, best_valid_epoch)))
if model_name in ['BiLSTM', 'CNN', 'Cha_CNN_LSTM']:
    for sentence, label in test_loader:

        outputs = model(Variable(sentence.view(1, -1)).cuda(),
            train=False)
        _, predicted = torch.max(outputs.data, 1) # 두번째 아웃풋 값은
            argmax 를 반환

        total += label.size(0) # batch 쓰는 경우.
        if (predicted.cpu().numpy() != label.numpy()):
            sentence = [ix_to_word[word_idx] for word_idx in
                sentence.long()[0][0].numpy()]
            missed_pairs.append('label: ' + str(label.numpy()[0][0])
                                + '\t' + 'predicted: ' + str(predicted.cpu().numpy()) +

```

```

        '\t' + ' '.join(sentence)+'\n')
    correct += (predicted.cpu() == label).sum()

elif model_name in ['Siamese_BiLSTM', 'Siamese_CNN']:
    for sentence_1, sentence_2, label in test_loader:

        output = model(Variable(sentence_1.view(1, -1)).cuda(),
                        Variable(sentence_2.view(1, -1)).cuda(), train=False)
        _, predicted = torch.max(output.data, 1) # 두번째 아웃풋 값은
        argmax 를 반환

        total += label.size(0) # batch 쓰는 경우.
        if (predicted.cpu().numpy() != label.numpy()):
            sentence_1 = [ix_to_word[word_idx] for word_idx in
                          sentence_1.long()[0][0].numpy()]
            sentence_2 = [ix_to_word[word_idx] for word_idx in
                          sentence_2.long()[0][0].numpy()]
            missed_pairs.append('label: ' + str(label.numpy()[0][0])
                               + '\t' + 'predicted: ' + str(predicted.cpu().numpy()) +
                               '\t' + ' '.join(sentence_1)+'\t'+
                               '\t'.join(sentence_2)+'\n')
        correct += (predicted.cpu() == label).sum()

print('Test Accuracy of the best valid accuracy model: %0.2f %%' %
      (100 * correct / total))
# write result
with
    open('./result/'+directory_name+'/%s_hid%d_D%0.2f_Acc%0.2f_best_valid_accuracy.txt'
          % (model_name, hidden_size, dropout_rate, 100 * correct /
             total), 'a', encoding='utf-8') as w:

        w.write('[setting]:
            '+'\tbatch_size\t'+str(batch_size)+'\ttemp_size\t'+str(embedding_size)+'\tHid\t'
        w.write('[Test Accuracy of the model]: %0.2f %% \n' % (100 *
            correct / total))
        w.write('[saved to]:
            './models/%s_hid%d_D%0.2f_Acc%0.2f_best_valid_accuracy.pkl\n'
            % (model_name, hidden_size, dropout_rate, 100 * correct /
               total))
        torch.save(model.state_dict(),
            './models/'+directory_name+'/%s_hid%d_D%0.2f_Acc%0.2f_best_valid_accuracy.pkl'
            % (model_name, hidden_size, dropout_rate, 100 * correct /
               total))

        for miss in missed_pairs:
            w.write(miss)
        w.write('-'*100+'\n\n')

```

Code 11: utils.py

```

import torch
import torch.nn as nn
import torch.nn.functional as F

from torch.utils.data import Dataset, DataLoader

```

```

from torch.autograd import Variable

import sys
import numpy
import os.path

import unicodedata
import string
import math

train_data_path = sys.argv[1] #"train_data.txt"
test_data_path = sys.argv[2] #"test_data.txt"

# 윈도우에서 스페인어 보기 위해 .
# Turn a Unicode string to plain ASCII, thanks to
# http://stackoverflow.com/a/518232/2809427
def unicodeToAscii(s):
    all_letters = string.ascii_letters + " .,;'"
    return ''.join(
        c for c in unicodedata.normalize('NFD', s)
        if unicodedata.category(c) != 'Mn'
        and c in all_letters
    )
#print(unicodeToAscii('Ślusàrski'))

def make_or_load_dict(train_data_path, character=False):
    something_to_ix = 'word_to_ix_' if not character else
        'character_to_ix_' # 캐릭터레벨은, 데이터 저장시에 캐릭터 단위로하면
        되도록 해둠.

    if not os.path.exists('./vocab'):
        os.makedirs('./vocab')

    # word_to_ix
    if os.path.isfile('./vocab/'+something_to_ix+train_data_path):
        word_to_ix = {}
        with open('./vocab/'+something_to_ix+train_data_path, 'r',
            encoding='utf-8') as dictionary:
            for line in dictionary.readlines():
                word, idx = line.strip().split('\t')
                word_to_ix[word] = int(idx)

    else:
        word_to_ix = {"unk":0, "endofsentence":1}
        with open(train_data_path, 'r', encoding='utf-8') as data:
            for line in data.readlines():
                sentence_1, sentence_2, label =
                    line.lower().strip().split('\t')
                word_list = (sentence_1+' '+sentence_2).split()
                for word in word_list:
                    if word not in word_to_ix:
                        word_to_ix[word] = len(word_to_ix)

        with open('./vocab/'+something_to_ix+train_data_path, 'w',
            encoding='utf-8') as w:
            for word, idx in word_to_ix.items():

```

```

        w.write(word+'\t'+str(idx)+'\n')

# ix to word
ix_to_word = [ i for i in range(len(word_to_ix))]
for word, idx in word_to_ix.items():
    ix_to_word[idx] = word

with
    open('./vocab/'+something_to_ix+train_data_path[:-4]+'_inverse_dict.txt',
        'w', encoding='utf-8') as w:
        for idx, word in enumerate(ix_to_word):
            w.write(word+'\t'+str(idx)+'\n')

vocab_size=len(word_to_ix)
print(vocab_size)
return word_to_ix, ix_to_word, vocab_size

def prepare_sequence(seq, to_ix):
    idxs = [to_ix[w] if w in to_ix else to_ix['unk'] for w in seq]
    #tensor = torch.FloatTensor([idxs])
    tensor = torch.LongTensor([idxs])
    return tensor

def prepare_label(label, to_ix):
    idxs = to_ix[label]
    tensor = torch.LongTensor([idxs])
    return tensor

# pad sequences from
https://docs.google.com/presentation/d/18JvZ5n49tt3-8nhBappz6aaraZMazF06v2YuQxGVK
def pad_sequences(vectorized_seqs, seq_lengths):
    seq_lengths_tensor = torch.LongTensor(seq_lengths)
    seq_tensor = torch.zeros((len(vectorized_seqs),
        seq_lengths_tensor.max())).long()
    for idx, (seq, seq_len) in enumerate(zip(vectorized_seqs,
        seq_lengths_tensor)):
        seq_tensor[idx, :seq_len] = torch.LongTensor(seq)
    return seq_tensor

class AlignmentDataset(Dataset):
    """ Alignment dataset."""
    # Initialize your data, download, etc.
    def __init__(self, data_path, word_to_ix, batch_size):
        data_file = open(data_path, 'r', encoding='utf-8')
        self.data = data_file.readlines()
        self.len = math.floor(len(self.data)/batch_size)
        self.word_to_ix = word_to_ix
        self.label_to_ix = {"1": 0, "2": 1}
        self.batch_size = batch_size
        #self.label_to_ix = {"Alignment": 0, "None": 1}

    def __getitem__(self, index):

```

```

vectorized_seqs, seq_lengths = [], []
label_batch = torch.zeros(self.batch_size, 1).long()

for i in range(self.batch_size):
    self.line = self.data[index*self.batch_size + i]
    sentence_1, sentence_2, label =
        self.line.lower().strip().split('\t')

    self.sentence = (sentence_1+' '+sentence_2).split()
    self.sentence_tensor = prepare_sequence(self.sentence,
        self.word_to_ix)
    seq_lengths.append(len(self.sentence))
    vectorized_seqs.append(self.sentence_tensor)

    self.label = prepare_label(label, self.label_to_ix)
    label_batch[i] = self.label

return pad_sequences(vectorized_seqs, seq_lengths),
    label_batch

def __len__(self):
    return self.len

class AlignmentDataset_seperate_sent(Dataset):
    """ Alignment dataset."""
    # Initialize your data, download, etc.
    def __init__(self, data_path, word_to_ix, batch_size):
        data_file = open(data_path, 'r', encoding='utf-8')
        self.data = data_file.readlines()
        self.len = math.floor(len(self.data)/batch_size)
        self.word_to_ix = word_to_ix
        self.label_to_ix = {"1": 0, "2": 1}
        self.batch_size = batch_size
        #self.label_to_ix = {"Alignment": 0, "None": 1}

    def __getitem__(self, index):
        vectorized_seqs_1, vectorized_seqs_2, seq_lengths_1,
            seq_lengths_2 = [], [], [], []
        label_batch = torch.zeros(self.batch_size, 1).long()

        for i in range(self.batch_size):
            self.line = self.data[index*self.batch_size + i]
            sentence_1, sentence_2, label =
                self.line.lower().strip().split('\t')

            self.sentence_1 = sentence_1.split()
            self.sentence_1_tensor = prepare_sequence(self.sentence_1,
                self.word_to_ix)
            seq_lengths_1.append(len(self.sentence_1))
            vectorized_seqs_1.append(self.sentence_1_tensor)

            self.sentence_2 = sentence_2.split()
            self.sentence_2_tensor = prepare_sequence(self.sentence_2,
                self.word_to_ix)
            seq_lengths_2.append(len(self.sentence_2))

```



```

        vectorized_seqs_2.append(self.sentence_2_tensor)

        self.label = prepare_label(label, self.label_to_ix)
        label_batch[i] = self.label

    return pad_sequences(vectorized_seqs_1, seq_lengths_1),
           pad_sequences(vectorized_seqs_2, seq_lengths_2),
           label_batch

def __len__(self):
    return self.len

class AlignmentDataset_cha_cnn(Dataset):
    """ Alignment dataset."""
    # Initialize your data, download, etc.
    def __init__(self, data_path, character_to_ix):
        data_file = open(data_path, 'r', encoding='utf-8')
        self.data = data_file.readlines()
        self.len = len(self.data)
        self.character_to_ix = character_to_ix
        self.label_to_ix = {"1": 0, "2": 1}
        #self.label_to_ix = {"Alignment": 0, "None": 1}

    def __getitem__(self, index):
        self.line = self.data[index]
        sentence_1, sentence_2, label =
            self.line.lower().strip().split('\t')

        self.sentence = (sentence_1+' '+sentence_2).split()
        self.sentence = [prepare_sequence(list(word),
            character_to_ix) for word in self.sentence]
        self.label = prepare_label(label, label_to_ix)

        return self.sentence, self.label

    def __len__(self):
        return self.len

class GenerateDataset(Dataset):
    """ Alignment dataset."""
    # Initialize your data, download, etc.
    def __init__(self, data_path, word_to_ix, batch_size):
        data_file = open(data_path, 'r', encoding='utf-8')
        self.data = data_file.readlines()
        self.len = math.floor(len(self.data)/batch_size)
        self.word_to_ix = word_to_ix
        self.batch_size = batch_size

    def __getitem__(self, index):
        vectorized_seqs_1, vectorized_seqs_2, seq_lengths_1,
            seq_lengths_2 = [], [], [], []

        for i in range(self.batch_size):
            self.line = self.data[index*self.batch_size + i]

```

```

        sentence_1, sentence_2, label =
            self.line.lower().strip().split('\t')

        self.sentence_1 = sentence_1.split()
        self.sentence_1_tensor = prepare_sequence(self.sentence_1,
            self.word_to_ix)
        seq_lengths_1.append(len(self.sentence_1))
        vectorized_seqs_1.append(self.sentence_1_tensor)

        self.sentence_2 = sentence_2.split()
        self.sentence_2_tensor = prepare_sequence(self.sentence_2,
            self.word_to_ix)
        seq_lengths_2.append(len(self.sentence_2))
        vectorized_seqs_2.append(self.sentence_2_tensor)

    return pad_sequences(vectorized_seqs_1, seq_lengths_1),
        pad_sequences(vectorized_seqs_2, seq_lengths_2)

def __len__(self):
    return self.len

class GenerateDataset_AE(Dataset):
    """ Alignment dataset."""
    # Initialize your data, download, etc.
    def __init__(self, data_path, word_to_ix, batch_size):
        data_file = open(data_path, 'r', encoding='utf-8')
        self.data = data_file.readlines()
        self.len = math.floor(len(self.data)/batch_size)
        self.word_to_ix = word_to_ix
        self.batch_size = batch_size

    def __getitem__(self, index):
        vectorized_seqs, seq_lengths = [], []

        for i in range(self.batch_size):
            self.line = self.data[index*self.batch_size + i]
            sentence_1, sentence_2, label =
                self.line.lower().strip().split('\t')

            self.sentence = (sentence_1+' '+sentence_2).split()
            self.sentence_tensor = prepare_sequence(self.sentence,
                self.word_to_ix)
            seq_lengths.append(len(self.sentence))
            vectorized_seqs.append(self.sentence_tensor)

        return pad_sequences(vectorized_seqs, seq_lengths),
            pad_sequences(vectorized_seqs, seq_lengths)

    def __len__(self):
        return self.len

class ContrastiveLoss(torch.nn.Module):
    """
    Borrowed from:
    https://github.com/harveyslash/Facial-Similarity-with-Siamese-Networks-in-PyTorch

```

```

Contrastive loss function.
Based on:
    http://yann.lecun.com/exdb/publis/pdf/hadsell-chopra-lecun-06.pdf
"""

def __init__(self, margin=2.0):
    super(ContrastiveLoss, self).__init__()
    self.margin = margin

def forward(self, output1, output2, label):
    euclidean_distance = F.pairwise_distance(output1, output2)
    loss_contrastive = torch.mean((1-label.float()) *
                                   torch.pow(euclidean_distance, 2) +
                                   (label.float()) *
                                   torch.pow(torch.clamp(self.margin -
                                                           euclidean_distance, min=0.0), 2))

    return loss_contrastive

# from
    http://pytorch.org/tutorials/intermediate/seq2seq\_translation\_tutorial.html
# This is a helper function to print time elapsed and estimated time
# remaining given the current time and progress %.
import time
import math

def asMinutes(s):
    m = math.floor(s / 60)
    s -= m * 60
    return '%dm %ds' % (m, s)

def timeSince(since, percent):
    now = time.time()
    s = now - since
    es = s / (percent)
    rs = es - s
    return '%s (- %s)' % (asMinutes(s), asMinutes(rs))

```

Code 12: models.py

```

import torch
import torch.nn as nn
import torch.nn.functional as F

from torch.autograd import Variable

import sys
import numpy
import math

# BiRNN Model (Many-to-One)
class BiLSTM(nn.Module):
    def __init__(self, vocab_size, embedding_size, hidden_size,

```

```

        num_layers, num_classes, dropout_rate):
    super(BiLSTM, self).__init__()
    self.hidden_size = hidden_size
    self.dropout = nn.Dropout(p=dropout_rate)

    self.word_embedding = nn.Embedding(vocab_size, embedding_size)

    self.num_layers = num_layers
    self.lstm = nn.LSTM(embedding_size, self.hidden_size,
                        num_layers, batch_first=True, bidirectional=True)
    self.fc = nn.Linear(self.hidden_size*2, num_classes) # 2 for
        bidirection

def forward(self, sentence, train = True):
    embeds = self.word_embedding(sentence)
    if train:
        embeds = self.dropout(embeds)

    # Set initial states
    h0 = Variable(torch.zeros(self.num_layers*2, embeds.size(0),
                            self.hidden_size)).cuda() # 2 for bidirection
    c0 = Variable(torch.zeros(self.num_layers*2, embeds.size(0),
                            self.hidden_size)).cuda()

    # Forward propagate RNN
    out, _ = self.lstm(embeds, (h0, c0))

    # Decode hidden state of last time step
    if train:
        out = self.fc(self.dropout(out[:, -1, :]))
    else:
        out = self.fc(out[:, -1, :])

    return out

class CNN(nn.Module):
    def __init__(self, vocab_size, embedding_size, num_classes,
                 dropout_rate, Ci = 1, kernel_num = 100, \
                 kernel_sizes=[3,4,5]):
        super(CNN, self).__init__()
        self.embed = nn.Embedding(vocab_size, embedding_size)
        self.padding = nn.ReflectionPad2d((0,0,1,1))
        self.convs1 = nn.ModuleList([nn.Conv2d(Ci, kernel_num, (K,
            embedding_size)) for K in kernel_sizes])
        self.dropout = nn.Dropout(dropout_rate)
        self.highway_t = nn.Linear(len(kernel_sizes)*kernel_num,
            len(kernel_sizes)*kernel_num) # square matrix
        self.highway_g = nn.Linear(len(kernel_sizes)*kernel_num,
            len(kernel_sizes)*kernel_num) # square matrix
        self.fc1 = nn.Linear(len(kernel_sizes)*kernel_num,
            num_classes)

def highway(self, input_, num_layers=1, bias=-2.0):
    """
    Recently, Kim Yoon's text CNN model usually use highway
    networks.

```

```

Highway Network (cf. http://arxiv.org/abs/1505.00387).
borrowed from https://github.com/mkroutikov/tf-lstm-char-cnn
t = sigmoid(Wy + b)
z = t * g(Wy + b) + (1 - t) * y
where g is nonlinearity, t is transform gate, and (1 - t) is
      carry gate.
"""
for idx in range(num_layers):
    t = F.sigmoid(self.highway_t(input_) + bias)
    g = F.relu(self.highway_g(input_))

    output = t * g + (1. - t) * input_
    input_ = output

return output

def forward(self, x, train = True):
    x = self.embed(x) # (N,W,D)

    x = x.unsqueeze(1) # (N,Ci,W,D) # N 은 배치수, Ci 가 채널수, W 가
    단어 윈도우수-3개보다 작으면 안됨 . D가 embedding_size
    x = self.padding(x)
    x = [F.relu(conv(x)).squeeze(3) for conv in self.convs1]
    #[(N,Co,W), ...]*len(Ks)
    x = [F.max_pool1d(i, i.size(2)).squeeze(2) for i in x]
    #[(N,Co), ...]*len(Ks)
    x = torch.cat(x, 1)

    x = self.highway(x, 1, 0)

    if train:
        x = self.dropout(x) # (N,len(Ks)*Co)
        logit = self.fc1(x) # (N,C)
    return logit

class Cha_CNN_LSTM(nn.Module):
    def __init__(self, vocab_size, embedding_size, num_classes,
        dropout_rate, Ci = 1, kernel_num = 100, \
        kernel_sizes=[3,4,5]):
        super(Cha_CNN_LSTM, self).__init__()
        self.embed = nn.Embedding(vocab_size, embedding_size)
        self.padding = nn.ReflectionPad2d((0,0,1,1))
        self.convs1 = nn.ModuleList([nn.Conv2d(Ci, kernel_num, (K,
            embedding_size)) for K in kernel_sizes])
        self.dropout = nn.Dropout(dropout_rate)
        self.fc1 = nn.Linear(len(kernel_sizes)*kernel_num,
            num_classes)

    def forward(self, x, train = True):
        x = self.embed(x) # (N,W,D)

        x = x.unsqueeze(1) # (N,Ci,W,D) # N 은 배치수, Ci 가 채널수, W 가
        단어 윈도우수-3개보다 작으면 안됨 . D가 embedding_size
        x = self.padding(x)
        x = [F.relu(conv(x)).squeeze(3) for conv in self.convs1]

```

```

        #[(N,Co,W), ...]*len(Ks)
    x = [F.max_pool1d(i, i.size(2)).squeeze(2) for i in x]
        #[(N,Co), ...]*len(Ks)
    x = torch.cat(x, 1)

    if train:
        x = self.dropout(x) # (N,len(Ks)*Co)
    logit = self.fcl(x) # (N,C)
    return logit

class Siamese_CNN(nn.Module):
    def __init__(self, vocab_size, num_classes, embedding_size=100,
        max_sequence_length = 30, kernel_size = 3, kernel_num=32):
        super(Siamese_CNN, self).__init__()
        self.max_sequence_length = max_sequence_length
        self.kernel_size = kernel_size

        #ABCNN 논문에서 BCNN 의 네트워크 구조를 빌림 .
        self.layer1 = nn.Sequential(
            nn.ReflectionPad2d((0,0,self.kernel_size-1,self.kernel_size-1)),
            nn.Conv2d(embedding_size, kernel_num, (self.kernel_size,
                1)),
            nn.ReLU(),
            nn.AvgPool2d((self.kernel_size,1), stride=1))
        self.layer2 = nn.Sequential(
            nn.ReflectionPad2d((0,0,self.kernel_size-1,self.kernel_size-1)),
            nn.Conv2d(kernel_num, kernel_num, (self.kernel_size, 1)),
            nn.ReLU(),
            nn.AvgPool2d((max_sequence_length+2,1), stride=1))

        self.LogisticRegression = nn.Linear(kernel_num*2, num_classes)

        self.embed = nn.Embedding(vocab_size, embedding_size)
        self.sentence_pad = nn.ReflectionPad2d((0,0,0,1))

    def siamese(self, x):
        x = self.embed(x) # (N,W,D)
        x = x.unsqueeze(1) # (N,Ci,W,D) # N 은 배치수, Ci 가 채널수, W 가
            단어 윈도우수-3개보다 작으면 안됨 . D가 embedding_size
        x = x.permute(0,3,2,1) # (N,D,W,Ci)
        if x.size(2)>self.max_sequence_length:
            x=x[:, :, :self.max_sequence_length, :]
        while (x.size(2)<self.max_sequence_length):
            x = self.sentence_pad(x)

        out = self.layer1(x)
        out = self.layer2(out)
        out = out.view(out.size(0), -1)

        return out

    def forward(self, sentence_1, sentence_2, train = True):
        represent_1 = self.siamese(sentence_1)
        represent_2 = self.siamese(sentence_2)
        similarity = torch.cat([represent_1, represent_2],1)

```

```

        logit = self.LogisticRegression(similarity)

    return logit

# Siamese_BiLSTM Model (Many-to-One)
class Siamese_BiLSTM(nn.Module):
    def __init__(self, vocab_size, embedding_size, hidden_size,
                  num_layers, num_classes, dropout_rate):
        super(Siamese_BiLSTM, self).__init__()
        self.hidden_size = hidden_size
        self.dropout = nn.Dropout(p=dropout_rate)
        self.word_embedding = nn.Embedding(vocab_size, embedding_size)

        self.num_layers = num_layers
        self.lstm = nn.LSTM(embedding_size, self.hidden_size,
                             num_layers, batch_first=True, bidirectional=True)

        self.LogisticRegression = nn.Linear(self.hidden_size*2*2,
                                              num_classes) # 2 for bidirection, 2 for siamese

    def siamese(self, sentence, train = True):
        embeds = self.word_embedding(sentence)
        if train:
            embeds = self.dropout(embeds)

        # Set initial states
        h0 = Variable(torch.zeros(self.num_layers*2, embeds.size(0),
                                   self.hidden_size)).cuda() # 2 for bidirection
        c0 = Variable(torch.zeros(self.num_layers*2, embeds.size(0),
                                   self.hidden_size)).cuda()

        # Forward propagate BRNN
        out, _ = self.lstm(embeds, (h0, c0))

        # Decode hidden state of last time step
        if train:
            out = self.dropout(out[:, -1, :])
        else:
            out = out[:, -1, :]

    return out

    def forward(self, sentence_1, sentence_2, train = True):
        represent_1 = self.siamese(sentence_1, train)
        represent_2 = self.siamese(sentence_2, train)
        similarity = torch.cat([represent_1, represent_2], 1)
        logit = self.LogisticRegression(similarity)

    return logit

```

Code 13: For hyper-parameter tuning. multi-train_text.sh

```
#!/bin/sh
```

```

echo -n "type [model name]: "
read model_name
echo model_name: $model_name

# echo -n "type [train data]: "
# read train_data
# echo train_data: $train_data

# echo -n "type [valid data]: "
# read valid_data
# echo valid_data: $valid_data

# echo -n "type [test data]: "
# read test_data
# echo test_data: $test_data

train_data=train_WD_0.10_aug_train_text.txt
valid_data=valid_WD_0.10_aug_train_text.txt
test_data=test_text.txt

echo -n "type [directory name]: "
read directory_name
echo directory_name: $directory_name

#embedding_size=300
#hidden_size=256
#dropout=0
for hidden_size in 32 64 128 256 512
do
    for dropout in 0 0.1 0.2 0.5
    do
        for try in {1..5..1}
        do
            echo hidden_size: $hidden_size
            echo dropout: $dropout
            python main.py $train_data $valid_data $test_data
                $directory_name $model_name $hidden_size $dropout
        done
    done
done
done

```

A.3 Generation of Aligned-conversation - Whole codes

Code 14: main_enc_dec.py

```

import torch
import torch.nn as nn
import torch.nn.functional as F

from torch.utils.data import Dataset, DataLoader
from torch.autograd import Variable

```



```

import sys
import numpy
import random

from utils import *
from models import *

# Hyper Parameters
max_sequence_length = 50
max_vocabulary_size = 25000
embedding_size = 300
hidden_size = int(sys.argv[6])#256
num_layers = 1
batch_size = 150
num_epochs = 50
learning_rate = 0.001
dropout_rate = float(sys.argv[7]) #0
teacher_forcing_ratio = float(sys.argv[8]) #0.5

train_data_path = sys.argv[1]
valid_data_path = sys.argv[2]
test_data_path = sys.argv[3] # test_twitter.txt
directory_name = sys.argv[4] #'twit_gen'

encoder_name = 'AttENC'
decoder_name = sys.argv[5] #'DEC'

word_to_ix, ix_to_word, vocab_size =
    make_or_load_dict(train_data_path, character=False)
num_classes = vocab_size

encoder = EncoderRNN(vocab_size, embedding_size, hidden_size,
    num_layers, 2, dropout_rate)
if decoder_name == 'AttDEC':
    decoder = AttnDecoderRNN(vocab_size, embedding_size,
        hidden_size, num_layers, num_classes, dropout_rate,
        max_sequence_length)
else:
    decoder = DecoderRNN(vocab_size, embedding_size, hidden_size,
        num_layers, num_classes, dropout_rate)

encoder = encoder.cuda()
decoder = decoder.cuda()
print(encoder)
print(decoder)

if decoder_name == 'AE': #for autoencoder
    train_dataset = GenerateDataset_AE(train_data_path, word_to_ix,
        batch_size)
    valid_dataset = GenerateDataset_AE(valid_data_path, word_to_ix,
        1)
    test_dataset = GenerateDataset_AE(test_data_path, word_to_ix, 1)
else:
    train_dataset = GenerateDataset(train_data_path, word_to_ix,
        batch_size)
    valid_dataset = GenerateDataset(valid_data_path, word_to_ix, 1)
    test_dataset = GenerateDataset(test_data_path, word_to_ix, 1)

```

```

train_loader = DataLoader(dataset=train_dataset, shuffle=True)
valid_loader = DataLoader(dataset=valid_dataset)
test_loader = DataLoader(dataset=test_dataset)
print('length of train dataset', len(train_dataset)*batch_size)
print('length of valid dataset', len(valid_dataset))
print('length of test dataset', len(test_dataset))

def train(input_variable, target_variable, batch_size, encoder,
          decoder, encoder_optimizer, decoder_optimizer, criterion,
          train_data = True):
    encoder_optimizer.zero_grad()
    decoder_optimizer.zero_grad()

    loss = 0

    encoder_outputs, encoder_hidden = encoder(input_variable)

    SOS_token = 0
    SOS_token_batch = torch.zeros(batch_size, 1).long()
    decoder_input = Variable(torch.LongTensor(SOS_token_batch))
    decoder_input = decoder_input.cuda()
    decoder_hidden = decoder.initHidden(batch_size, encoder_hidden)

    decoder_input_list = torch.unbind(target_variable, dim=1)

    # Choose whether to use teacher forcing
    use_teacher_forcing = random.random() < teacher_forcing_ratio
    if use_teacher_forcing:
        if decoder_name == 'AttDEC':
            for next_input in decoder_input_list:
                decoder_output, decoder_hidden, decoder_attention =
                    decoder(decoder_input, decoder_hidden,
                           encoder_outputs)
                loss += criterion(decoder_output, next_input)
                decoder_input = next_input.unsqueeze(1)
        else:
            for next_input in decoder_input_list:
                decoder_output, decoder_hidden = decoder(decoder_input,
                                                         decoder_hidden)
                loss += criterion(decoder_output, next_input)
                decoder_input = next_input.unsqueeze(1)
    else:
        if decoder_name == 'AttDEC':
            for next_input in decoder_input_list:
                decoder_output, decoder_hidden, decoder_attention =
                    decoder(decoder_input, decoder_hidden,
                           encoder_outputs)
                loss += criterion(decoder_output, next_input)

                # Get most likely word index (highest value) from output
                topv, topi = decoder_output.data.topk(1)
                decoder_input = Variable(topi).cuda() # Chosen word is
                next_input

        else:

```

```

        for next_input in decoder_input_list:
            decoder_output, decoder_hidden = decoder(decoder_input,
                                                    decoder_hidden)
            loss += criterion(decoder_output, next_input)

            # Get most likely word index (highest value) from output
            topv, topi = decoder_output.data.topk(1)
            decoder_input = topi.cuda() # Chosen word is next input

    if train_data:
        loss.backward()
        torch.nn.utils.clip_grad_norm(encoder.parameters(), 1)
        torch.nn.utils.clip_grad_norm(decoder.parameters(), 1)
        encoder_optimizer.step()
        decoder_optimizer.step()

    return loss.data[0] / target_variable.size()[0]

def evaluate(encoder, decoder, input_variable):
    encoder_outputs, encoder_hidden = encoder(input_variable,
                                              train=False)
    #max_sequence_length = encoder_outputs.size(1) # for
    #attention-simple dot product mode

    SOS_token = 0
    decoder_input = Variable(torch.LongTensor([[SOS_token]]))
    decoder_input = decoder_input.cuda()
    decoder_hidden = encoder_hidden
    decoder_hidden = decoder.initHidden(1, encoder_hidden)

    decoded_words = []
    decoder_attentions = torch.zeros(max_sequence_length,
                                     max_sequence_length)

    stop_at_next_endofsentence = 0
    if decoder_name == 'AttDEC':
        for di in range(max_sequence_length):
            decoder_output, decoder_hidden, decoder_attention =
                decoder(decoder_input, decoder_hidden,
                        encoder_outputs, train = False)
            decoder_attentions[di] = decoder_attention.view(-1,
                                                            max_sequence_length).data

            # Choose top word from output
            topv, topi = decoder_output.data.topk(1)
            ni = topi[0][0]
            decoded_words.append(ni)
            if ni ==1: break

            # Next input is chosen word
            decoder_input = Variable(torch.LongTensor([[ni]])).cuda()
    else:
        for _ in range(max_sequence_length):
            decoder_output, decoder_hidden = decoder(decoder_input,
                                                    decoder_hidden, train = False)

```

```

        # Choose top word from output
        topv, topi = decoder_output.data.topk(1)
        ni = topi[0][0]
        decoded_words.append(ni)
        if ni ==1:
            if decoder_name == 'AE' and stop_at_next_endofsentence
                !=1:
                    stop_at_next_endofsentence = 1
            elif decoder_name == 'AE' and
                stop_at_next_endofsentence == 1:
                    break
            else:
                break

        # Next input is chosen word
        decoder_input = Variable(torch.LongTensor([[ni]]).cuda())

    if decoder_name == 'AttDEC':
        return decoded_words, decoder_attentions[:di + 1]
    else:
        return decoded_words

print_every = (len(train_dataset)/2)

# Loss and Optimizer
loss_function = nn.CrossEntropyLoss() # = log softmax + NLL loss
encoder_optimizer = torch.optim.Adam(encoder.parameters(),
    lr=learning_rate)
decoder_optimizer = torch.optim.Adam(decoder.parameters(),
    lr=learning_rate)

print_loss_total = 0
keep_valid_loss = 1e06
if not os.path.exists('./models'):
    os.makedirs('./models')
if not os.path.exists('./models/'+directory_name):
    os.makedirs('./models/'+directory_name)

for epoch in range(num_epochs):
    for i, (sentence_1, sentence_2) in enumerate(train_loader):
        start = time.time()
        sentence_1 = Variable(sentence_1.view(batch_size, -1)).cuda()
        sentence_2 = Variable(sentence_2.view(batch_size, -1)).cuda()
        loss = train(sentence_1, sentence_2, batch_size, encoder,
            decoder, encoder_optimizer, decoder_optimizer,
            loss_function)
        print_loss_total += loss

    if ((i+1)+epoch*len(train_dataset)) % print_every == 0:
        print_loss_avg = print_loss_total / print_every
        print_loss_total = 0

        # valid set test
        valid_loss = 0
        for sentence_1, sentence_2 in valid_loader:
            sentence_1 = Variable(sentence_1.view(1, -1)).cuda()

```

```

sentence_2 = Variable(sentence_2.view(1, -1)).cuda()
loss = train(sentence_1, sentence_2, 1, encoder,
              decoder, encoder_optimizer, decoder_optimizer,
              loss_function, train_data = False)
valid_loss += loss

if valid_loss < keep_valid_loss:
    keep_valid_loss = valid_loss
    best_valid_epoch = epoch + 1
    torch.save(encoder.state_dict(),
               './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_epoch%d_best_v'
               % (encoder_name, hidden_size, dropout_rate,
                  teacher_forcing_ratio, (epoch+1)))
    torch.save(decoder.state_dict(),
               './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_epoch%d_best_v'
               % (decoder_name, hidden_size, dropout_rate,
                  teacher_forcing_ratio, (epoch+1)))

    print('%s (%d %d%%) loss %.4f, valid loss: %.4f <best'
          ' valid>' % (timeSince(start,
                                   ((i+1)+epoch*len(train_dataset)) /
                                   (num_epochs*len(train_dataset))),
                      (i+1)+epoch*len(train_dataset),
                      ((i+1)+epoch*len(train_dataset))
                      / (num_epochs*len(train_dataset))
                      * 100, print_loss_avg,
                      valid_loss))

else:
    torch.save(encoder.state_dict(),
               './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_epoch%d.pkl'
               % (encoder_name, hidden_size, dropout_rate,
                  teacher_forcing_ratio, (epoch+1)))
    torch.save(decoder.state_dict(),
               './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_epoch%d.pkl'
               % (decoder_name, hidden_size, dropout_rate,
                  teacher_forcing_ratio, (epoch+1)))
    print('%s (%d %d%%) loss %.4f, valid loss: %.4f' %
          (timeSince(start, ((i+1)+epoch*len(train_dataset)) /
                          (num_epochs*len(train_dataset))),
           (i+1)+epoch*len(train_dataset),
           ((i+1)+epoch*len(train_dataset))
           / (num_epochs*len(train_dataset))
           * 100, print_loss_avg,
           valid_loss))

losses = 0

# Test the Model
if not os.path.exists('./result'):
    os.makedirs('./result')
if not os.path.exists('./result/'+directory_name):
    os.makedirs('./result/'+directory_name)

encoder.load_state_dict(torch.load('./models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_epoch%d.pkl'
                                   % (encoder_name, hidden_size, dropout_rate,
                                      teacher_forcing_ratio)))

```

```

decoder.load_state_dict(torch.load('./models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%
    % (decoder_name, hidden_size, dropout_rate,
    teacher_forcing_ratio)))

with
    open('./result/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_generation.txt'%(deco
    hidden_size, dropout_rate, teacher_forcing_ratio), 'a',
    encoding='utf-8') as w:

    attention_result = []
    for sentence_1, sentence_2 in test_loader:
        sentence_variable = Variable(sentence_1.view(1, -1)).cuda()

        if decoder_name == 'AttDEC':
            predicted, attention = evaluate(encoder, decoder,
                sentence_variable)
            attention_result.append(attention)
        else:
            predicted = evaluate(encoder, decoder, sentence_variable)

        sentence_1 = [ix_to_word[word_idx] for word_idx in
            sentence_1.long()[0][0].numpy()]
        sentence_2 = [ix_to_word[word_idx] for word_idx in
            sentence_2.long()[0][0].numpy()]
        predicted = [ix_to_word[word_idx] for word_idx in predicted]
        w.write('Input_sentence: \t' + ' '.join(sentence_1)+'\n')
        w.write('Target_sentence: \t' + ' '.join(sentence_2)+'\n')
        w.write('Genera_sentence: \t' + ' '.join(predicted)+'\n')

    w.write('[setting]:
        '+'\tbatch_size\t'+str(batch_size)+'\temb_size\t'+str(embedding_size)+'\tHid\t
    w.write('saved to
        './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f.pkl\n'
        % (encoder_name, hidden_size, dropout_rate,
        teacher_forcing_ratio))
    w.write('saved to
        './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f.pkl\n'
        % (decoder_name, hidden_size, dropout_rate,
        teacher_forcing_ratio))
    torch.save(encoder.state_dict(),
        './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_last_epoch.pkl'
        % (encoder_name, hidden_size, dropout_rate,
        teacher_forcing_ratio))
    torch.save(decoder.state_dict(),
        './models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_last_epoch.pkl'
        % (decoder_name, hidden_size, dropout_rate,
        teacher_forcing_ratio))
    w.write('-'*100+'\n\n')

# load the best valid model.
#best_valid_epoch = 5
encoder.load_state_dict(torch.load('./models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%
    % (encoder_name, hidden_size, dropout_rate,
    teacher_forcing_ratio, best_valid_epoch)))
decoder.load_state_dict(torch.load('./models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%
    % (decoder_name, hidden_size, dropout_rate,

```

```

teacher_forcing_ratio, best_valid_epoch)))

with
    open('./result/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_with_best_valid_loss.%s' %
         (hidden_size, dropout_rate, teacher_forcing_ratio), 'a',
         encoding='utf-8') as w:

    attention_result = []
    for sentence_1, sentence_2 in test_loader:
        sentence_variable = Variable(sentence_1.view(1, -1)).cuda()

        if decoder_name == 'AttDEC':
            predicted, attention = evaluate(encoder, decoder,
                                           sentence_variable)
            attention_result.append(attention)
        else:
            predicted = evaluate(encoder, decoder, sentence_variable)

    sentence_1 = [ix_to_word[word_idx] for word_idx in
                  sentence_1.long()[0][0].numpy()]
    sentence_2 = [ix_to_word[word_idx] for word_idx in
                  sentence_2.long()[0][0].numpy()]
    predicted = [ix_to_word[word_idx] for word_idx in predicted]
    w.write('Input_sentence: \t' + ' '.join(sentence_1)+'\n')
    w.write('Target_sentence: \t' + ' '.join(sentence_2)+'\n')
    w.write('Genera_sentence: \t' + ' '.join(predicted)+'\n')

w.write('[setting]:
'+'\tbatch_size\t'+str(batch_size)+'\temb_size\t'+str(embedding_size)+'\tHid\t'+str(hidden_size)+'\n')
w.write('saved to
./models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_best_valid_loss.pkl\n' %
      (encoder_name, hidden_size, dropout_rate,
       teacher_forcing_ratio))
w.write('saved to
./models/'+directory_name+'/%s_hid%d_D%0.2f_tfr%0.1f_best_valid_loss.pkl\n' %
      (decoder_name, hidden_size, dropout_rate,
       teacher_forcing_ratio))
w.write('-'*100+'\n\n')

```

Code 15: Encoder (BiLSTM)

```

class EncoderRNN(nn.Module):
    def __init__(self, vocab_size, embedding_size, hidden_size,
                 num_layers, num_classes, dropout_rate):
        super(EncoderRNN, self).__init__()
        self.hidden_size = hidden_size
        self.dropout = nn.Dropout(p=dropout_rate)
        self.word_embedding = nn.Embedding(vocab_size, embedding_size)

        self.num_layers = num_layers
        self.lstm = nn.LSTM(embedding_size, self.hidden_size,
                             num_layers, batch_first=True, bidirectional=True)
        self.fc = nn.Linear(self.hidden_size*2, num_classes) # *2 for
        bidirection, BiLSTM 을 오토인코더로 초기화할 때 필요.

```

```

def initHidden(self, batch_size):
    # Set initial states
    h0 = Variable(torch.zeros(self.num_layers*2, batch_size,
                              self.hidden_size)).cuda() # 2 for bidirection
    c0 = Variable(torch.zeros(self.num_layers*2, batch_size,
                              self.hidden_size)).cuda()
    return (h0, c0)

def forward(self, sentence, train = True):
    #print(sentence)
    embeds = self.word_embedding(sentence)
    if train:
        embeds = self.dropout(embeds)

    # Set initial states
    h0 = Variable(torch.zeros(self.num_layers*2, embeds.size(0),
                              self.hidden_size)).cuda() # 2 for bidirection
    c0 = Variable(torch.zeros(self.num_layers*2, embeds.size(0),
                              self.hidden_size)).cuda()

    # Forward propagate RNN
    output, hidden = self.lstm(embeds, (h0, c0))

    return output, hidden

```

Code 16: Decoder (LSTM)

```

class DecoderRNN(nn.Module):
    def __init__(self, vocab_size, embedding_size, hidden_size,
                  num_layers, num_classes, dropout_rate):
        super(DecoderRNN, self).__init__()
        self.hidden_size = hidden_size
        self.dropout = nn.Dropout(p=dropout_rate)

        self.word_embedding = nn.Embedding(vocab_size, embedding_size)
        self.to_decoder_init_hidden = nn.Linear(2*num_layers,
                                                  num_layers)

        self.num_layers = num_layers
        self.lstm = nn.LSTM(embedding_size, self.hidden_size,
                              num_layers, batch_first=True)
        self.fc = nn.Linear(self.hidden_size, num_classes) # 2 for
        bidirection

    def initHidden(self, batch_size, init_hidden):
        # Set initial states
        decoder_init_hidden =
            self.to_decoder_init_hidden(init_hidden[0].transpose(2,
                0)).transpose(2, 0).contiguous()
        decoder_init_c = Variable(torch.zeros(self.num_layers,
            batch_size, self.hidden_size)).cuda()
        return (decoder_init_hidden, decoder_init_c)

```



```

def forward(self, sentence, hidden, train = True):
    embeds = self.word_embedding(sentence)
    if train:
        embeds = self.dropout(embeds)

    # Forward propagate RNN
    #embeds = F.relu(embeds)
    output, hidden = self.lstm(embeds, hidden)

    # Decode hidden state of last time step
    if train:
        output = self.fc(self.dropout(output[:, -1, :]))
    else:
        output = self.fc(output[:, -1, :])

    return output, hidden

```

Code 17: Attention-Decoder (LSTM with attention)

```

# from pytorch tutorial
http://pytorch.org/tutorials/intermediate/seq2seq\_translation\_tutorial.html
class AttnDecoderRNN(nn.Module):
    def __init__(self, vocab_size, embedding_size, hidden_size,
        num_layers, num_classes, dropout_rate,
        max_sequence_length=50):
        super(AttnDecoderRNN, self).__init__()
        self.num_layers = num_layers
        self.hidden_size = hidden_size
        self.dropout = nn.Dropout(p=dropout_rate)
        self.max_sequence_length = max_sequence_length

        self.W_A = nn.Linear(embedding_size + 2*self.hidden_size,
            self.max_sequence_length)
        self.word_embedding = nn.Embedding(vocab_size, embedding_size)
        self.W_C = nn.Linear(embedding_size + 2*self.hidden_size,
            2*self.hidden_size)
        self.lstm = nn.LSTM(2*self.hidden_size, 2*self.hidden_size,
            num_layers, batch_first=True)
        self.fc = nn.Linear(2*self.hidden_size, num_classes)

    def initHidden(self, batch_size, init_hidden):
        # Set initial states
        tmp = torch.unbind(init_hidden[0], dim=0)
        decoder_init_hidden = torch.cat(tmp, dim=1).unsqueeze(0)
        decoder_init_c = Variable(torch.zeros(self.num_layers,
            batch_size, 2*self.hidden_size)).cuda()
        return (decoder_init_hidden, decoder_init_c)

    def forward(self, input, hidden, encoder_outputs, train = True):
        embeds = self.word_embedding(input)
        if train:
            embeds = self.dropout(embeds)

```

```

encoder_outputs_max = Variable(torch.zeros( embeds.size(0),
      self.max_sequence_length, 2*self.hidden_size)).cuda()
encoder_outputs_max[:, :encoder_outputs.size(1), :] =
    encoder_outputs

score = self.W_A(torch.cat((embeds,
    hidden[0].transpose(0,1)), 2))
attn_weights = F.softmax(score, dim=2)
context = torch.bmm(attn_weights, encoder_outputs_max)

output = torch.cat((embeds, context), 2)
output = F.relu(self.W_C(output))

output, hidden = self.lstm(output, hidden)

# Decode hidden state of last time step
if train:
    output = self.fc(self.dropout(output[:, -1, :]))
else:
    output = self.fc(output[:, -1, :])

return output, hidden, attn_weights

# from Effective Approaches to Attention-based Neural Machine
# Translation https://arxiv.org/pdf/1508.04025.pdf.
class AttnDecoderRNN_simple_dotproduct(nn.Module):
    def __init__(self, vocab_size, embedding_size, hidden_size,
        num_layers, num_classes, dropout_rate, max_sequence_length):
        super(AttnDecoderRNN_simple_dotproduct, self).__init__()
        self.num_layers = num_layers
        self.hidden_size = hidden_size
        self.dropout = nn.Dropout(p=dropout_rate)

        self.word_embedding = nn.Embedding(vocab_size, embedding_size)
        self.W_C = nn.Linear(embedding_size + 2*self.hidden_size,
            2*self.hidden_size)
        self.lstm = nn.LSTM(2*self.hidden_size, 2*self.hidden_size,
            num_layers, batch_first=True)
        self.fc = nn.Linear(2*self.hidden_size, num_classes)

    def initHidden(self, batch_size, init_hidden):
        # Set initial states
        tmp = torch.unbind(init_hidden[0], dim=0)
        decoder_init_hidden = torch.cat(tmp, dim=1).unsqueeze(0)
        decoder_init_c = Variable(torch.zeros(self.num_layers,
            batch_size, 2*self.hidden_size)).cuda()
        return (decoder_init_hidden, decoder_init_c)

    def forward(self, input, hidden, encoder_outputs, train = True):
        embeds = self.word_embedding(input)
        if train:
            embeds = self.dropout(embeds)

```

```

W_enc= encoder_outputs # for simple dot product attention
score = torch.bmm(hidden[0].transpose(0,1),
    W_enc.transpose(1,2)) # eq(7), dot product
attn_weights = F.softmax(score, dim=2)
context = torch.bmm(attn_weights, encoder_outputs)

output = torch.cat((embeds, context), 2)
output = F.tanh(self.W_C(output)) # eq(5)
#output = F.relu(self.W_C(output)) # eq(5)

output, hidden = self.lstm(output, hidden)

# Decode hidden state of last time step
if train:
    output = self.fc(self.dropout(output[:, -1, :])) # eq (6)
else:
    output = self.fc(output[:, -1, :])
return output, hidden, attn_weights

```

A.4 Marker Validation - Whole codes

Code 18: find_marker.py

```

import torch
import torch.nn as nn
import torch.nn.functional as F

from torch.utils.data import Dataset, DataLoader
from torch.autograd import Variable

import sys
import numpy

from utils import *
from models import *

# Hyper Parameters
max_sequence_length = 30
max_vocabulary_size = 25000
embedding_size = 300
hidden_size = int(sys.argv[5]) # 256
num_layers = 2
num_classes = 2
batch_size = 1
num_epochs = 30
learning_rate = 0.001
dropout_rate = float(sys.argv[6]) #0

train_data_path = 'train_WD_0.10_aug_train_twitter.txt'
unlabeled_data_path = sys.argv[1] # preprocessed_50000unlabeled.txt
directory_name = sys.argv[2] #'twit_marker'
accuracy = float(sys.argv[3]) # 77.00

```

```

model_name = sys.argv[4] #'CNN'

word_to_ix, ix_to_word, vocab_size =
    make_or_load_dict(train_data_path, character=False)

def model(x):
    return {
        'BiLSTM': BiLSTM(vocab_size, embedding_size, hidden_size,
                           num_layers, num_classes, dropout_rate),
        'CNN': CNN(vocab_size, embedding_size, num_classes,
                   dropout_rate, kernel_num=hidden_size),
        'Cha_CNN_LSTM': Cha_CNN_LSTM(vocab_size, embedding_size,
                                       num_classes, dropout_rate, kernel_num=hidden_size),
        'Siamese_BiLSTM': Siamese_BiLSTM(vocab_size, embedding_size,
                                           hidden_size, num_layers, num_classes, dropout_rate),
        'Siamese_CNN': Siamese_CNN(vocab_size, num_classes,
                                    embedding_size, kernel_num=hidden_size),
    }.get(x)

model = model(model_name)
model.cuda()
model.load_state_dict(torch.load('./models/'+directory_name+'/%s_hid%d_D%0.2f_Acc%0.2f.txt'%
                                (model_name, hidden_size, dropout_rate, accuracy)))
#model.load_state_dict(torch.load('./models/'+directory_name+'/%s_hid%d_D%0.2f_Acc%0.2f.txt'%
                                (model_name, hidden_size, dropout_rate, accuracy)))
print(model)

# Predict whether tweet-reply pair is alignment or not
correct = 0
total = 0
if not os.path.exists('./result'):
    os.makedirs('./result')
if not os.path.exists('./result/'+directory_name):
    os.makedirs('./result/'+directory_name)

with open(unlabeled_data_path, 'r', encoding = 'utf-8') as tweet, \
    open('./result/'+directory_name+'/Align_%s_hid%d_D%0.2f_Acc%0.2f.txt'%(model_name,
                                   hidden_size, dropout_rate, accuracy), 'w', encoding
        ='utf-8') as Align, \
    open('./result/'+directory_name+'/None_%s_hid%d_D%0.2f_Acc%0.2f.txt'%(model_name,
                                   hidden_size, dropout_rate, accuracy), 'w', encoding
        ='utf-8') as none:

    count = 0
    for sentences in tweet.readlines():
        count += 1
        if count % 100 == 0:
            print('pair_count', count)

        sentence_1_origin, sentence_2_origin, _ =
            sentences.lower().strip().split('\t')

        if model_name in ['BiLSTM', 'CNN', 'Cha_CNN_LSTM']:
            sentence_concat = (sentence_1_origin+
                              ' '+sentence_2_origin).split()
            sentence_concat = prepare_sequence(sentence_concat,
                                                word_to_ix)

```

```

        outputs = model(Variable(sentence_concat.view(batch_size,
            -1)).cuda(), train=False)
        _, predicted = torch.max(outputs.data, 1) # 두번째 아웃풋
            값은 argmax 를 반환

    elif model_name in ['Siamese_BiLSTM', 'Siamese_CNN']:
        sentence_1 = sentence_1_origin.split()
        sentence_1 = prepare_sequence(sentence_1, word_to_ix)

        sentence_2 = sentence_2_origin.split()
        sentence_2 = prepare_sequence(sentence_2, word_to_ix)

        output = model(Variable(sentence_1.view(batch_size,
            -1)).cuda(), Variable(sentence_2.view(batch_size,
            -1)).cuda(), train=False)
        _, predicted = torch.max(output.data, 1) # 두번째 아웃풋 값은
            argmax 를 반환

    if predicted.cpu().numpy() == 0:
        Align.write(sentence_1_origin+'\t'+sentence_2_origin+'\t'+ '1'+ '\n')
            # alignment pair 를 원문 그대로 저장.
    else:
        none.write(sentence_1_origin+'\t'+sentence_2_origin+'\t'+ '2'+ '\n')
            # non-alignment pair 를 원문 그대로 저장.

# 코퍼스의 각 단어별 빈도를 센 뒤, 빈도순으로 나열해서 저장.
# from
    https://www.kaggle.com/anokas/data-analysis-xgboost-starter-0-35460-1b
from collections import Counter
counts = Counter()
align_count = 0
with
    open('./result/'+directory_name+'/Align_%s_hid%D%0.2f_Acc%0.2f.txt'%(model_name,
        hidden_size, dropout_rate, accuracy), 'r', encoding = 'utf-8')
        as r_align, \
    open('./result/'+directory_name+'/s_Align_marker%0.2f.txt'%(model_name,
        accuracy), 'w', encoding = 'utf-8') as w_align:
        for line in r_align.readlines():
            counts.update(line.rstrip().split())
            align_count += 1

weights = {word: count/align_count for word, count in
    counts.items()}
print('align_count', align_count)
for key, count in sorted(weights.items(), key=lambda x: x[1],
    reverse=True):
    w_align.write(key+'\t'+str(count)+'\n')

counts_none = Counter()
none_count = 0
with
    open('./result/'+directory_name+'/None_%s_hid%D%0.2f_Acc%0.2f.txt'%(model_name,
        hidden_size, dropout_rate, accuracy), 'r', encoding = 'utf-8')

```

```

as r_none, \
open('./result/'+directory_name+'/%s_None_marker_%0.2f.txt'%(model_name,
    accuracy), 'w', encoding = 'utf-8') as w_none:
for line in r_none.readlines():
    counts_none.update(line.rstrip().split())
    none_count +=1

weights_none = {word: count/none_count for word, count in
    counts_none.items()}
print('none_count', none_count)

for key, count in sorted(weights_none.items(), key=lambda x:
    x[1], reverse=True):
    w_none.write(key+'\t'+str(count)+'\n')

```

A.5 Spanish Conversation Pairs

Each example is stored in the format; [sentence_1 / sentence_2 / label (1=alignment, 2=non-alignment)]

A.5.1 Textbook corpus

Table A1: Examples of labeled data: Textbook (1)

¡Hola! ¿Qué tal?	¡Hola!	1
¿Cómo te llamas?	Me llamo María Pérez.	1
¿Y tú? ¿Cuál es tu nombre?	Mi nombre es Pedro García.	1
Encantada de conocerte.	Igualmente.	1
María, ¿cuál es el correo electrónico del profesor López?	A ver. ... ¡Ah, sí! Su correo es lopez@gmail.com.	1
A ver. ... ¡Ah, sí! Su correo es lopez@gmail.com.	¿Cómo se escribe?	1
¿Cómo se escribe?	Se escribe L-O-P-E-Z.	1
¡Muchas gracias!	¡De nada!	1
¡Hasta otro día!	¡Vale! ¡Hasta pronto!	1
¡Vale! ¡Hasta pronto!	¡Adiós!	1
Buenos días, profesor Moreno. ¿Cómo está usted?	Muy bien, gracias. ¿Y usted?	1
Muy bien, gracias. ¿Y usted?	Bien, gracias. ¡Hasta luego!	1
Bien, gracias. ¡Hasta luego!	¡Adiós!	1
Buenas tardes, señora.	Buenas tardes, doña María. ¿Cómo va todo?	1
Buenas tardes, doña María. ¿Cómo va todo?	No muy bien.	1
No muy bien.	¿Qué pasa?	1
¿Qué pasa?	Estoy cansada.	1
¡Hola, Paco! Buenas noches. ¿Qué tal?	Bien.	1
¿Y tú?	Así, así.	1
¡Chao!	¡Chao! ¡Hasta pronto!	1
Muchas gracias, señor.	De nada, señorita.	1
¡Adiós!	¡Hasta la vista!	1
¿Cómo te llamas?	Me llamo Esther.	1
Me llamo Esther.	¿Y tu apellido?	1
¿Y tu apellido?	Kim.	1
¿Cómo se escribe?	K de kilo, I de iglesia, M de Madrid.	1
¿Cómo se llama usted?	Mi nombre es Raúl.	1
¿Cómo se escribe su nombre?	R-A-U-L.	1
Este es mi amigo Juan.	¡Mucho gusto!	1
¡Mucho gusto!	¡Igualmente!	1
¡Mire! Le presento a la profesora Claudia.	¡Encantado!	1
¡Encantado!	¡Encantada!	1
¡Hola, Sumi!	¡Hola, Carlos!	1
¿Qué tal?	Muy bien, gracias.	1
¿Y tú? ¿Cómo estás?	¡Genial!	1
¿Qué tal tu clase de español?	¡Excelente!	1
En mi clase hay muchos estudiantes inteligentes.	¡Qué interesante!	1
¿Hay aire acondicionado en tu aula?	¡Por supuesto! También hay un ordenador nuevo y un proyector muy moderno, una pizarra verde muy larga y muchas tizas de colores.	1
¡Por supuesto! También hay un ordenador nuevo y un proyector muy moderno, una pizarra verde muy larga y muchas tizas de colores.	¿De qué colores?	1
¿De qué colores?	Azules, blancas, rojas y amarillas.	1

Table A2: Examples of labeled data: Textbook (2)

¡Sumi! ¿Qué grande es tu mochila! ¿Qué hay?	Hay un cuaderno verde, un cuaderno de color naranja, lápices, bolígrafos y un ordenador portátil nuevo.	1
Hay un cuaderno verde, un cuaderno de color naranja, lápices, bolígrafos y un ordenador portátil nuevo.	¡Muchas cosas!	1
Bueno, hasta pronto, Sumi.	¡Hasta otro día, Carlos!	1
¿Qué hay en el salón de clase?	En el salón de clase hay una pizarra, unas sillas y unos libros.	1
¿Qué hay en el salón de clase?	En el aula no hay reloj.	1
¿Qué hay en el aula?	En el salón de clase hay una pizarra, unas sillas y unos libros.	1
¿Qué hay en el aula?	En la clase no hay reloj.	1
¿Hay una computadora?	Sí, hay una computadora.	1
¿Hay una computadora?	Sí, hay un ordenador.	1
¿Hay una computadora?	No, no hay computadora.	1
¿Hay una computadora?	No, no hay ordenador.	1
¿Hay un ordenador?	Sí, hay una computadora.	1
¿Hay un ordenador?	Sí, hay un ordenador.	1
¿Hay un ordenador?	No, no hay computadora.	1
¿Hay un ordenador?	No, no hay ordenador.	1
¿De qué color es el libro?	Es rojo.	1
¿De qué color es el libro?	Es de color naranja.	1
¡Hola! Buenos días, Sofia.	¡Buenos días!	1
Mira, te presento a mi mejor amigo.	¡Hola!	1
¡Hola!	¿Qué tal? Perdona, ¿cómo te llamas?	1
¿Qué tal? Perdona, ¿cómo te llamas?	Mi nombre es Enrique, pero mis amigos me llaman Quique.	1
Mi nombre es Enrique, pero mis amigos me llaman Quique.	Mucho gusto.	1
Mucho gusto.	Encantado.	1
¿Eres español?	No, no soy español. Soy de México.	1
No, no soy español. Soy de México.	¿De qué parte de México eres?	1
¿De qué parte de México eres?	Soy de México, D.F. (Distrito Federal). ¿Y tú?	1
Soy de México, D.F. (Distrito Federal). ¿Y tú?	Yo soy de Sevilla. Soy española.	1
Buenos días.	Buenos días.	1
¿Su nombre, por favor?	Soy Luis Carlos González Pérez.	1
¿Cuál es su profesión?	Soy ingeniero electrónico.	1
¿Cuántos idiomas habla?	Hablo tres idiomas: español, francés e inglés.	1
¿Cuáles son sus cualidades?	Soy trabajador, sincero, entusiasta, amable y puntual.	1
¿Cuál es su nacionalidad?	Soy chileno.	1
Muchas gracias por su tiempo, Sr. Pérez.	A usted por la oportunidad.	1
¿Eres Juan?	No, no soy Juan. Soy Pedro.	1
¿Es usted profesor?	No, no soy profesor. Soy estudiante.	1
¿Usted es profesor?	No, no soy profesor. Soy estudiante.	1
Tú eres coreano, ¿no?	Sí, soy coreano.	1
Usted es de Argentina, ¿verdad?	No, no soy de Argentina. Soy mexicana.	1
¿Cuál es tu nombre?	Mi nombre es Juan.	1

Table A3: Examples of labeled data: Textbook (3)

¿Cuáles son tus apellidos?	García López.	1
¿Quién es usted?	Soy Pedro, el profesor de español.	1
¿Quiénes son ellos?	Son mis alumnos.	1
¿De quién es este libro?	Es de Mario.	1
¿Cuántos estudiantes hay en la clase?	Hay once estudiantes.	1
¿Cuánto vale la entrada?	Dos euros.	1
¿Dónde viven ustedes?	Vivimos en Seúl, la capital de Corea del Sur.	1
¿Cuándo es tu cumpleaños?	Es el quince de agosto.	1
¿Por qué estás triste?	Porque no tengo amigos.	1
¿Cómo estás?	Estoy muy bien.	1
¿Cómo es tu profesor?	Es alto y simpático.	1
¿De dónde es usted?	Soy de Japón.	1
¿Con quién vive Juan?	Vive con su familia.	1
¿De dónde eres?	Soy de México.	1
¿De dónde sois vosotras?	Somos de España.	1
¿De dónde es ella?	Es de los Estados Unidos.	1
¿De dónde es usted?	Soy de Corea del Sur.	1
¿De dónde eres?	Soy coreano. Soy de Seúl.	1
¿Y usted? ¿De dónde es?	Soy española. Soy de Madrid.	1
¿De dónde son ustedes?	Yo soy japonesa. Jean y Marie son franceses y Silvia es inglesa.	1
¿Qué lenguas habla usted?	Hablo ruso, francés e inglés.	1
¿Qué lengua hablan en Brasil?	Hablan portugués.	1
¿Cuál es la nacionalidad de Juan?	Su nacionalidad es colombiana.	1
¡Bienvenida a Madrid, Mison!	Muchas gracias, señora Pérez.	1
¿Qué tal el viaje?	Bien, pero estoy un poco cansada.	1
Es un viaje bastante largo.	¡Sí, es verdad!	1
Mison, esta es tu habitación. El cuarto de baño está a la derecha de tu habitación y el comedor está al lado de la cocina.	Muy amable, señora.	1
Una pregunta. ¿Hay una farmacia cerca de la casa?	¡Claro que sí! Hay una farmacia muy grande enfrente de la casa.	1
Y la universidad, ¿dónde está?	La universidad está un poco lejos, pero hay una parada del bus universitario delante de la casa.	1
Una última pregunta. ¿Hay un banco por aquí?	Hay muchos bancos dentro del centro comercial del barrio. El centro comercial está entre la oficina de correos y la plaza.	1
¡Qué bien! Mil gracias, señora Pérez.	¡De nada, Mison!	1
¿Cómo es tu profesor?	Es simpático e idealista.	1
¿Cómo estás? ¿Estás enferma?	Estoy un poco cansada, pero estoy bien.	1
¿Está soltero tu hermano?	No, está casado.	1
¡Oye, estás muy guapa hoy!	Gracias. Tú también estás muy bonita.	1
¿Qué hay en la habitación?	Hay una cama.	1
¿Qué hay en la mesa?	Hay unos libros.	1
¿Dónde estás?	Estoy en la cocina.	1
¿Dónde está Alemania?	Está en Europa.	1
¿Dónde está el Museo Nacional?	Está en la Plaza Mayor, al lado del ayuntamiento.	1

Table A4: Examples of labeled data: Textbook (4)

¿Está cerca de aquí?	Sí, sí, muy cerca.	1
¿Dónde está Corea?	Está en el noreste de Asia. Al oeste de Japón.	1
¿Cuál es la capital de Corea?	Es Seúl. Está en el centro de la península coreana.	1
Profesora Moreno, ¿tiene usted un coche nuevo?	Sí, tengo un Mercedes negro.	1
Esteban, ¿tienes clase ahora?	Sí, tengo clase de español.	1
José, ¿tienes tiempo?	Lo siento, no tengo tiempo.	1
¿Cuántos años tiene usted?	Tengo cuarenta y un años.	1
¿Qué edad tiene usted?	Tengo cuarenta y un años.	1
¿Qué te pasa? ¿Estás mal?	Tengo frío.	1
¿Qué te pasa? ¿Estás mal?	Tengo calor.	1
¿Qué te pasa? ¿Estás mal?	Tengo prisa.	1
¿Qué te pasa? ¿Estás mal?	Tengo hambre.	1
¿Qué te pasa? ¿Estás mal?	Tengo sed.	1
¿Qué te pasa? ¿Estás mal?	Tengo miedo.	1
¿Qué te pasa? ¿Estás mal?	Tengo sueño.	1
¿Qué te pasa? ¿Estás mal?	Tengo fiebre.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho frío.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho calor.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho prisa.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho hambre.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho sed.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho miedo.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho sueño.	1
¿Qué te pasa? ¿Estás mal?	Tengo mucho fiebre.	1
¿Tienes dolor de cabeza?	No, tengo dolor de estómago.	1
Eres nueva en la universidad, ¿verdad?	Sí, soy estudiante de intercambio de Corea.	1
Sí, soy estudiante de intercambio de Corea.	¿Qué estudias?	1
¿Qué estudias?	Estudio administración de empresas.	1
Estudio administración de empresas.	¡Qué coincidencia! Yo también estudio la misma carrera.	1
¡Qué coincidencia! Yo también estudio la misma carrera.	¿Sí? Jajaja. ¡Increíble!	1
¿Cómo te llamas?	Pablo, ¿y tú?	1
Pablo, ¿y tú?	Me llamo Mison.	1
¿Dónde vives, Mison?	Vivo en el centro de la ciudad en una pensión familiar.	1
¡Qué bien hablas español!	Gracias, pero no tanto.	1
Estos días, veo programas televisivos y escucho la radio para practicar el idioma.	¡Qué diligente!	1
Yo solo hablo español.	Bueno, desde ahora, yo te enseño coreano y tú me enseñas español.	1
Bueno, desde ahora, yo te enseño coreano y tú me enseñas español.	¡Buena idea!	1
¿El señor Pérez trabaja en una cafetería?	No, trabaja en una librería.	1
¿Hablan ustedes francés?	Sí, hablamos francés y coreano.	1
¿Aprende usted chino?	Sí, aprendo chino y japonés.	1

Table A5: Examples of labeled data: Textbook (5)

¿Aquí venden billetes?	Sí, vendemos billetes.	1
¿Qué bebes, tequila o vino?	Yo no bebo.	1
¿Dónde vive usted?	Vivo en la calle Libertad.	1
¿Qué escribes?	Escribo una carta a mi amigo.	1
¿Qué tal, María?	¡Hola, Carlos!	1
¿Estás ocupada este fin de semana?	Sí, mañana es el cumpleaños de mi abuelo.	1
¿Cuántos años cumple tu abuelo?	Setenta y cinco años. Mi abuela también tiene la misma edad.	1
¿Dónde viven tus abuelos?	Ellos viven en Córdoba.	1
¿Viven solos tus abuelos?	No, ellos viven con mis tíos.	1
Juan, ¿qué es eso?	Es una foto de mi familia.	1
Es una foto de mi familia.	A ver ...	1
Mira, estos del centro son mis abuelos y a su derecha están mis padres. Esta es mi hermana Carolina y este soy yo.	¡Tienes una familia preciosa!	1
¡Tienes una familia preciosa!	Gracias.	1
¿Qué hacen tus padres?	Mi padre es médico y trabaja en un hospital muy grande y mi madre es profesora. ¿Y tus padres?	1
Mi padre es médico y trabaja en un hospital muy grande y mi madre es profesora. ¿Y tus padres?	Mi padre es diplomático y mi madre es ama de casa.	1
¿Vives con ellos?	No, yo vivo en la residencia universitaria porque ellos viven fuera del país.	1
¿Son bonitas estas flores?	Sí, son muy bonitas.	1
¿Cuánto cuesta ese bolígrafo?	Este bolígrafo cuesta diecinueve euros.	1
Aquellas casas son muy modernas, ¿verdad?	Sí, además tienen muchos árboles.	1
¿Qué es esto?	Es un ordenador portátil.	1
¿Cuánto vale aquello?	Vale cien euros.	1
¿Con quién compartes el piso?	Con un amigo mío.	1
¿Estas son mis llaves?	No, las tuyas están al lado del florero.	1
¿Dónde están mis zapatos?	Los tuyos están allí y los míos, aquí.	1
¿Es de Carmen este libro?	No, es mío. El suyo está en su escritorio.	1
¿De quién es la llave?	Es mía.	1
¿Este coche es tuyo?	No, no es mío. Es de Clara.	1
¿Son vuestros estos zapatos?	Sí, son nuestros.	1
El viernes es el cumpleaños de Carla.	¿Sí?	1
Pensamos hacer una "fiesta sorpresa" en su casa. ¿Pueden asistir?	¡Claro! Carla es una buena compañera.	1
¿A qué hora empieza la fiesta?	A las seis de la tarde.	1
A las seis de la tarde.	¡Estupendo!	1
¿Qué regalo piensas comprar?	Yo un perfume o una billetera.	1
¿Y tú?	No sé todavía.	1
Bueno, entonces nos vemos pasado mañana en el club. De allí vamos todos juntos a casa de Carla.	Sí, sí. Gracias.	1
¿Quién es?	¡Sorpresa! ¡Feliz cumpleaños, Carla!	1
¡Sorpresa! ¡Feliz cumpleaños, Carla!	¡Oh! ¡Gracias! ¡Muchas gracias a todos!	1
Adelante, por favor.	Gracias ...	1
Carla, estos son nuestros regalos.	¡Qué emoción! Sois muy buenos amigos.	1
¡Ahora la tarta y las velas! Uno, dos, tres... Cumpleaños feliz, cumpleaños feliz, te deseamos todos, cumpleaños feliz ¡Bravo! (aplausos)	¡Gracias, amigos!	1

Table A6: Examples of labeled data: Textbook (6)

Perdón, señor. ¿Qué hora tiene?	Es la una en punto.	1
Perdón, señor. ¿Qué hora tiene?	Son las tres y cuarto.	1
Perdón, señor. ¿Qué hora tiene?	Son las cinco y media.	1
Perdón, señor. ¿Qué hora tiene?	Son las seis menos cinco.	1
Perdón, señor. ¿Qué hora tiene?	Son casi las ocho.	1
Perdón, señor. ¿Qué hora tiene?	Es mediodía.	1
Perdón, señor. ¿Qué hora tiene?	Es medianoche.	1
¿A qué hora empieza la clase de español?	Empieza a las diez de la mañana y termina a las once y media.	1
¿A qué hora abren los bancos?	Abren a las nueve.	1
¿A qué hora cierran?	Cierran a las 4 de la tarde.	1
¿Trabajas por la mañana?	Sí, trabajo de nueve a doce.	1
Sí, trabajo de nueve a doce.	Pues, yo trabajo desde las diez hasta las dos de la tarde.	1
¿La oficina de Álvaro está en el séptimo piso?	No, está en la octava planta.	1
¿A qué piso va usted?	Al segundo.	1
¿Cuál es tu dirección?	Calle Gaya, número 35, cuarto B.	1
¿Qué día es hoy?	Hoy es jueves.	1
¿Qué días tienes la clase de español?	Los lunes y miércoles.	1
¿Qué haces los sábados?	Los sábados estoy en la pensión todo el día.	1
¿Qué fecha es hoy?	Hoy es (el) 11 de mayo.	1
¿Cuándo es tu cumpleaños?	Es el 2 de marzo.	1
Entonces, ¿a cuántos estamos?	Estamos a primero de marzo.	1
¡Buenas tardes!	¡Buenas tardes, señora.	1
¿Qué desea?	Quiero un paquete turístico para dos personas.	1
Quiero un paquete turístico para dos personas.	Muy bien.	1
¿Adónde quiere viajar?	No sé todavía. Como es un viaje para mis padres, quiero elegir un buen lugar, sobre todo, un lugar con buen clima.	1
¿Qué le parece Sudamérica?	¿Sabe qué tiempo hace allí?	1
¿Sabe qué tiempo hace allí?	El tiempo depende de cada país.	1
¿Qué tal el Caribe? Tenemos un crucero muy cómodo y elegante.	¿Sí? ¿Cómo es el clima allí?	1
¿Sí? ¿Cómo es el clima allí?	En el Caribe siempre hace buen tiempo, a veces hace viento, pero es muy suave y refrescante.	1
Es un lugar ideal para descansar.	¡Perfecto! Entonces, quiero un viaje al Caribe.	1
¿Qué le das a tu amiga?	Le doy un regalo.	1
¿Ves algo?	No, no veo nada sin mis gafas.	1
¿Cuántas estaciones hay en Corea?	Hay cuatro estaciones: primavera, verano, otoño e invierno.	1
¿Qué tiempo hace en verano en Corea?	Hace mucho calor y llueve mucho. El clima es muy húmedo.	1
¿Y en invierno?	Hace mucho frío. A veces nieva y, en general, el clima es seco.	1
¿Cómo es el clima de España en verano?	Hace mucho sol y hace muchísimo calor.	1
¿Usted tiene que ir a la biblioteca hoy?	Sí, tengo que estudiar porque tenemos un examen mañana.	1
Para la salud hay que hacer ejercicio.	¡Es verdad! También hay que dejar de fumar.	1
¿Qué vas a hacer este fin de semana?	Nada especial. Voy a estar en casa.	1
Nada especial. Voy a estar en casa.	Pues, ¿vamos al cine?	1

Table A7: Examples of labeled data: Textbook (7)

Pues, ¿vamos al cine?	¡Magnífico! ¿Qué película vamos a ver?	1
¡Magnífico! ¿Qué película vamos a ver?	Una película de acción.	1
Este sábado estrenan una película muy interesante.	¡Ah! ¿Sí?	1
¿En qué cine la van a proyectar?	En el cine que está enfrente del ayuntamiento.	1
¿A qué hora y dónde quedamos?	A las cinco de la tarde enfrente de la taquilla.	1
Primero, voy a ver el horario y te envío un mensaje a tu móvil. Luego, voy a reservar las entradas por Internet.	¡Perfecto!	1
Después de la película, ¿qué vamos a hacer?	¿Por qué no vamos a tomar una copa?	1
¿Por qué no vamos a tomar una copa?	¡Buena idea! Hasta el sábado.	1
¿Adónde van ustedes ahora?	Vamos a la fiesta de cumpleaños de Rosa.	1
¿Qué van a hacer ustedes este fin de semana?	Vamos a jugar al tenis. ¿Y usted?	1
Vamos a jugar al tenis. ¿Y usted?	Voy a visitar a mis padres.	1
¿Qué van a hacer los alumnos durante las vacaciones de invierno?	Van a viajar por países latinoamericanos.	1
¿Para quién es este regalo?	Es para ti.	1
¿Qué dicen tus padres de mí?	Hablan muy bien de ti.	1
¿Quieres venir a una fiesta conmigo?	Claro que sí. Contigo voy al fin del mundo.	1
¿Tiene María coche?	Sí, lo tiene.	1
¿Tienen ustedes los pasaportes?	No, no los tenemos. Los dejamos en casa.	1
Juan, ¿me quieres?	Claro. Te quiero mucho.	1
¿Conoces a Marta?	Sí, la conozco. Somos amigas.	1
Voy a casarme.	No lo creo.	1
¿Ves a aquel chico?	Sí, lo veo. / Sí, le veo.	1
¿Tomas un café?	No, no lo tomo por la noche.	1
¿Quién prepara el almuerzo en tu casa?	Mi madre lo prepara.	1
¿Comes pescado para cenar?	Sí, lo como todos los días.	1
¿Prefieres las novelas hispanoamericanas?	Sí, las prefiero.	1
¿Dónde compras la ropa?	La compro en un centro comercial.	1
¿Bebes mucho tequila?	No, no la bebo tanto.	1
¿Qué le vas a decir a José?	Le voy a decir que no puedo venir mañana.	1
¿Cuándo me das el libro?	Te lo doy mañana. Te lo prometo.	1
¿Me pasas las páginas amarillas?	Sí, te las paso.	1
¿A quién le escribes el correo electrónico?	Se lo escribo a mi jefe.	1
¿A quién le regalas estas flores tan bonitas?	Se las regalo a mi mujer.	1
¿Es verdad que María le quiere regalar su coche a su hermana?	Creo que sí. Se lo quiere regalar.	1
¿Es verdad que María le quiere regalar su coche a su hermana?	Creo que sí. Quiere regalárselo.	1
Hola, Pedro. ¿Cómo va todo?	¡Feliz!	1
Acabamos de ganar el partido de fútbol.	¡Enhorabuena!	1
¿Juegas todos los días al fútbol?	Sí, me encanta jugar al fútbol.	1
¿Y, a ti, qué deporte te gusta practicar?	A mí me gusta jugar al baloncesto, pero también me gusta ver el fútbol.	1
Me emociona ver los partidos de la Copa Mundial.	Sí, son emocionantes.	1
Aparte del fútbol, ¿qué otras cosas te gustan?	Bueno, me fascina escribir poemas y también tocar el piano.	1

Table A8: Examples of labeled data: Textbook (8)

¿Sí, diga?	Buenos días. Soy Clara Ruiz.	1
¿Puedo hablar con Susana?	Un momento por favor. Ahora se pone.	1
Hola, Clara. ¿Qué hay?	Hola, Susana. Mañana es el día de la excursión y hoy tenemos que hacer las compras.	1
¿Qué les gusta comer a los chicos de tu grupo?	Bueno, a algunos les gusta el jamón, a otros les gustan las aceitunas y a las chicas les gustan las frutas.	1
Bueno, a algunos les gusta el jamón, a otros les gustan las aceitunas y a las chicas les gustan las frutas.	Bien, entonces, voy a comprar jamón, aceitunas y frutas.	1
¿Algo más?	No, nada más.	1
Clara, también tienes que comprar bebidas.	Sí, sí. De acuerdo.	1
¿Sí, diga?	Hola, buenas tardes. ¿Está Ignacio?	1
Hola, buenas tardes. ¿Está Ignacio?	¿De parte de quién?	1
¿De parte de quién?	De José.	1
De José.	Un momento. Ahora se pone.	1
¿Bueno?	Hola, buenos días. ¿Puedo hablar con el profesor Moreno?	1
Hola, buenos días. ¿Puedo hablar con el profesor Moreno?	Sí, en seguida. ¿Quién habla?	1
Sí, en seguida. ¿Quién habla?	Habla Ana María.	1
Habla Ana María.	Un momentito, por favor.	1
¿Dígame?	¿Podría hablar con el señor Montero?	1
¿Podría hablar con el señor Montero?	Equivocado. ¿Qué número ha marcado?	1
Equivocado. ¿Qué número ha marcado?	Perdone.	1
¿Diga?	¿Está María?	1
¿Está María?	No, no está. ¿Quiere dejar algún recado?	1
No, no está. ¿Quiere dejar algún recado?	No, gracias. Vuelvo a llamarla.	1
¿A tus padres les gusta ver la televisión?	Sí, les encanta.	1
¿Le gusta al profesor Lee la comida italiana?	No lo sé. Pero a mí me gusta mucho la comida italiana.	1
Me gustan los perros.	A mí también.	1
Me gustan los perros.	A mí no.	1
No me gustan los gatos.	A mí tampoco.	1
No me gustan los gatos.	A mí sí.	1
¿Te gusta beber vino?	No, no me gusta.	1
No, no me gusta.	A mí tampoco.	1
¿Te gusta beber vino?	Sí, me gusta mucho.	1
Sí, me gusta mucho.	A mí también.	1
¿Qué te gusta desayunar?	Me gusta desayunar huevos revueltos con una tostada.	1
¿Qué te gusta comer?	Me gusta comer hamburguesa con patatas fritas.	1
¿Qué te gusta almorzar?	Me gusta almorzar hamburguesa con patatas fritas.	1
¿Qué te gusta cenar?	Me gusta cenar ensalada.	1
¿Sabes nadar?	No, no sé nadar.	1
¿Sabes que Juan viene a Corea este domingo?	Sí, lo sé.	1
¿Conoces al embajador de España en Corea?	Sí, lo conozco. Somos amigos.	1
¿Conoces Buenos Aires?	Sí, conozco muy bien la capital de Argentina.	1
¿Sabes esquiar?	Sí, sé esquiar muy bien.	1

Table A9: Examples of labeled data: Textbook (9)

¿Sabes esquiar?	No, no sé esquiar.	1
¡Buenas tardes! ¿Puedo ayudarla?	Sí, por favor.	1
Quiero comprar una falda.	¿Qué tipo de falda quiere?	1
¿Qué tipo de falda quiere?	Una falda corta.	1
¿De qué estilo?	Una falda vaquera de algodón.	1
¿De qué color?	Quiero una falda de color azul claro.	1
¿De qué talla?	No sé... una falda de talla mediana, o un poco más pequeña.	1
Tiene que probársela antes de comprarla. Cada marca tiene tallas diferentes.	¡Es verdad! También, según el diseño, a veces me va bien una talla más pequeña.	1
¿La quiere con bolsillos?	No, es mejor sin bolsillos.	1
No, es mejor sin bolsillos.	¡Ya la tengo!	1
Es muy bonita. Es la más bonita de todas.	¿Cuánto cuesta?	1
¿Cuánto cuesta?	A ver... Esta falda cuesta treinta euros.	1
Y está en oferta.	¡Qué bien! Entonces, la compro.	1
¿Qué desea?	Solo estoy mirando	1
¿Qué quiere comprar?	Solo estoy mirando	1
¿En qué puedo servirle?	Solo estoy mirando	1
¿Cuánto cuesta el vestido?	Cuesta cien euros.	1
¿Cuánto vale?	Cuesta cien euros.	1
En total, ¿cuánto es?	Son 200 euros.	1
¿Qué le parece esta falda?	Me queda muy bien.	1
¿Qué le parece esta falda?	Me queda un poco corta.	1
¿Puedo probarme estos pantalones?	Sí, claro. ¿Qué talla tiene usted?	1
¿Quiere pagar en efectivo o con tarjeta?	En efectivo.	1
¡Mario, despiértate! ¡Es la hora de levantarse! ¡Ya son las siete de la mañana! ¡Vas a llegar tarde!	¡Ya voy mamá!...	1
Con una taza de leche caliente y una cucharada de miel se te pasa.	Está bien, mamá.	1
Ahora mismo me levanto y me visto.	Primero, debes ducharte y no tardes.	1
Siempre pasas mucho tiempo peinándote y afeitándote.	¡Mamá! ¡Solo paso treinta minutos en el baño!	1
¡Date prisa! Tienes que desayunar y no te queda mucho tiempo.	¡No te preocupes, mamá! Tengo tiempo suficiente.	1
No olvides que, después de desayunar, debes cepillarte los dientes.	¡Ya lo sé mamá! ¡Eso puedo hacerlo en un segundo!	1
Te aseguro que voy a llegar a tiempo a la universidad.	Bueno, confío en ti.	1
¿Qué te duele?	Me duele la espalda.	1
¿Qué le duele a Ud.?	Me duele el estómago.	1
¿Qué le duele a ese hombre?	Le duelen las piernas.	1
¿Qué estás haciendo?	Estoy leyendo el periódico en el salón.	1
¿Estás estudiando la lección 12?	Sí, estoy estudiándola.	1
¿Estás estudiando la lección 12?	Sí, la estoy estudiando.	1
¿Maria está poniéndose la falda?	Sí, se la está poniendo.	1
¿Maria está poniéndose la falda?	Sí, está poniéndose la falda.	1
¿Qué haces?	Veo la televisión.	1
¿Qué haces?	Estoy viendo la televisión.	1

Table A10: Examples of labeled data: Textbook (10)

¿Qué estás haciendo?	Veo la televisión.	1
¿Qué estás haciendo?	Estoy viendo la televisión.	1
Estas vacaciones quiero ir de viaje por España. ¿Has estado en España alguna vez?	Sí, acabo de regresar de Zaragoza.	1
Sí, acabo de regresar de Zaragoza.	¡Qué suerte!	1
¿Qué lugares has visitado?	He recorrido casi todas las ciudades.	1
He recorrido casi todas las ciudades.	Entonces, ¿qué ciudades me recomiendas?	1
Entonces, ¿qué ciudades me recomiendas?	Yo creo que Madrid y Barcelona son las ciudades que debes visitar.	1
Y ¿has probado la comida española?	¡Por supuesto! He comido paella a la valenciana, gazpacho y diferentes tipos de tapas.	1
¡La comida española es muy sabrosa!	¡Ya veo!	1
¿Has visto algún espectáculo?	Bueno, en Sevilla he visto una corrida de toros.	1
Bueno, en Sevilla he visto una corrida de toros.	¡Qué emoción!	1
Estoy pensando usar el AVE para recorrer España. ¿Qué opinas?	¡Buena idea! ¡Buen viaje!	1
¿Ha salido el tren para Madrid?	No, aún	1
¿Ha salido el tren para Madrid?	todavía no ha salido.	1
¿Ha salido el tren para Madrid?	Sí, ya ha salido.	1
He perdido mi pasaporte.	¡Qué lástima!	1
¿Has estado alguna vez en Argentina?	Sí, dos veces.	1
¿Has estado alguna vez en Argentina?	No, no he estado nunca.	1
¿Has estado alguna vez en Argentina?	No, nunca he estado allí.	1
Hablas español muy bien. ¿Cuánto tiempo has estudiado español?	He estudiado español durante seis meses.	1
¿Habéis comprado manzanas?	Sí, ya las hemos comprado.	1
¿Habéis comprado manzanas?	No, todavía no las hemos comprado.	1
¿Están abiertos los bancos?	No, a esta hora están cerrados.	1
¿Hay algún taller cerca de aquí?	Sí, pero está cerrado ahora.	1
¿Qué tal?	Estoy muy cansado.	1
En España, ¿a qué hora se abren los bancos?	En España, los bancos se abren a las nueve y media de la mañana.	1
¿Hay algo en esa caja?	No, no hay nada en la caja.	1
¿Conoces a alguien en este hospital?	No, no conozco a nadie.	1
¿Vive alguien en esta casa?	No, no vive nadie.	1
¿Conoces a algún profesor?	No, no conozco a ninguno.	1
Si tiene alguna duda, puede llamarme inmediatamente.	No, no tengo ninguna. Gracias por su amabilidad.	1
¿Quiere algo más?	No, no quiero nada más.	1
¡Hola! ¡Buenas tardes!	¡Buenas tardes!	1
¿A qué hora sale el próximo tren a Sevilla?	Hay un AVE que sale a las dieciséis quince.	1
Hay un AVE que sale a las dieciséis quince.	¿Es el último?	1
¿Es el último?	No. El último sale a las veinte quince.	1
No. El último sale a las veinte quince.	¡Perfecto!	1
Quiero sacar dos billetes de ida y vuelta para las dieciséis quince.	Aquí tiene.	1
¿En qué andén está el tren?	En el siete.	1
¡Hola, Diana! ¿Qué tal el fin de semana? ¿Qué hiciste el sábado?	¡Hola! Por la tarde, jugué al tenis con mis amigos. Después, fuimos a un bar y hablamos mucho. Lo pasamos genial.	1

Table A11: Examples of labeled data: Textbook (11)

¿Y tú?	Me encontré con mi novia y comimos tapas en la Plaza Mayor. Después vimos una película.	1
Y ¿qué hiciste el domingo?	El domingo dormí muchas horas y almorcé con mis padres en un restaurante italiano. Por la tarde, salí con mi hermana y comimos en un parque.	1
El domingo dormí muchas horas y almorcé con mis padres en un restaurante italiano. Por la tarde, salí con mi hermana y comimos en un parque.	¿Qué bien!	1
En cambio, yo llamé a María y pasamos casi todo el día preparando una presentación en grupo para la clase.	¿Y no hicisteis nada más?	1
¿Y no hicisteis nada más?	Bueno, más tarde fuimos a la Puerta del Sol.	1
Y comimos churros con chocolate.	¿Son deliciosos!	1
¿Son deliciosos!	¡Así es!	1
¿A qué hora saliste ayer de la oficina?	Salí a las cinco de la tarde.	1
¿Qué pasó en el año 1492?	Cristóbal Colón llegó a América.	1
¿Qué tal Rafael?	¡Hola! ¿Qué tal?	1
Pareces contenta.	Sí, mucho.	1
La semana pasada fui de viaje a Andalucía y lo pasé muy bien.	¿Ah, sí?	1
¿Adónde fuiste?	Fuimos a la playa, nos bañamos y tomamos el sol durante varias horas.	1
Fuimos a la playa, nos bañamos y tomamos el sol durante varias horas.	¿Pasaron todo el día solo en la playa?	1
¿Pasaron todo el día solo en la playa?	¡Claro que no! Por la tarde fuimos al pueblo de Federico García Lorca. Allí visitamos su casa natal.	1
¿Y por la noche qué hicieron?	Fuimos a Granada y cenamos en un restaurante árabe.	1
¿Visitaron La Alhambra?	¡Claro! El domingo temprano visitamos La Alhambra.	1
Inmediatamente después, regresamos a Madrid.	¡Vaya! Eso sí fue un buen fin de semana.	1
¿Y tú adónde fuiste?	No hice nada en especial, pero la próxima semana voy a ir a Valencia.	1
¿Por qué no viniste a la fiesta ayer?	Porque tuve un accidente de tráfico.	1
¿Dónde pusiste la llave?	La puse encima de la mesa del comedor.	1
¿Qué dijiste?	No dije nada.	1
¿Fuiste a la biblioteca ayer?	No. Fui a casa de Juan a estudiar.	1
¿Quién fue tu profesor de español el año pasado?	Fue la profesora Torres.	1
¿Qué regalo te dieron en tu compañía la Navidad pasada?	Me dieron una botella de vino.	1
¿Qué pidieron anoche ustedes en el restaurante?	Juan pidió pollo asado con patatas y yo pedí pescado con verduras.	1
¿Durmieron ustedes bien anoche?	Carlos durmió bien, pero yo no dormí nada.	1
¿Cuánto tiempo hace que vives en España?	Hace cinco meses que vivo en España.	1
¿Cuánto tiempo hace que estudias español?	Hace seis meses que estudio español.	1
¿Cuánto tiempo hace que estudias español?	Estudio español desde hace seis meses.	1
¿Cuánto tiempo hace que te casaste?	Hace 20 años que me casé.	1
¿Cuánto tiempo hace que te casaste?	Me casé hace 20 años.	1
¿Cuánto tiempo hace que viniste a Corea?	Hace casi diez años que vine a Corea.	1
¿Cuánto tiempo hace que viniste a Corea?	Vine a Corea hace casi diez años.	1
¿Tienes familia en Toledo?	Sí, ¿no lo sabías? Soy toledana y de niña vivía allí.	1
¿Dónde vivías tú cuando eras pequeño?	Yo vivía en el centro de Madrid. Mi casa estaba cerca del Parque del Retiro.	1
Mi casa estaba cerca del Parque del Retiro.	¿Sí?	1
Y ¿cómo eras de niño? ¿Qué hacías?	No me acuerdo mucho, pero creo que era muy alegre. Me gustaba mucho jugar y pasaba los fines de semana jugando con mis amigos en el parque y montando en bicicleta.	1
¿Y tú cómo eras?	Yo era muy tímida y tranquila.	1
Pasaba mucho tiempo en casa con mi madre: veía la televisión, jugaba con mis muñecas y leía muchos cuentos de hadas.	¡Ahora comprendo por qué te gusta la lectura!	1

Table A12: Examples of labeled data: Textbook (12)

¿Hacías algún deporte?	No mucho, solo saltaba a la comba mientras mi madre contaba hasta cien.	1
¿Y tú?	Mi padre practicaba deportes acuáticos y durante las vacaciones solía llevarme al mar para nadar y pescar.	1
¿Qué bien recuerdas tu infancia!	¡Es verdad! Ahora creo que los recuerdos de la infancia son inolvidables.	1
¿Qué hacían los niños?	Los niños jugaban al fútbol en la calle.	1
¿Qué hacías cuando llegué?	Cuando llegaste, leía un libro.	1
¿Qué estabas haciendo cuando te llamé?	Estaba durmiendo en mi habitación.	1
¿Qué estabas haciendo cuando te llamé?	Estaba duchándome.	1
¿A qué hora salió el tren ayer?	No sé. Cuando llegué a la estación, el tren ya había salido.	1
¿Leíste el libro que te había recomendado?	Sí, lo leí.	1
¡Hola, chicos! ¿Qué hacen por acá?	Estamos hablando de las vacaciones.	1
Estamos hablando de las vacaciones.	¡Por fin! Mañana empiezan nuestras vacaciones de verano.	1
¡Por fin! Mañana empiezan nuestras vacaciones de verano.	¿Qué vas a hacer estas vacaciones?	1
¿Qué vas a hacer estas vacaciones?	Iré en avión a Cancún y pasaré allá un fin de semana fantástico. Me alojaré en un hotel muy bonito frente al mar y, por la noche, beberé y bailaré en alguna discoteca.	1
¿Bailas bien?	Soy ecuatoriana y nací bailando ...	1
Para mí, estas serán mis últimas vacaciones de soltera, así que debo aprovecharlas antes de volver a Ecuador.	¿Qué bien! ¡Qué buena idea!	1
¿Irás al concierto la próxima semana?	Sí, iré con mis amigos.	1
¿A qué hora terminará la clase?	Terminará a las cuatro de la tarde.	1
¿Qué hora es?	No sé. Serán las diez y media.	1
¿Habrá mucho tráfico en el centro de la ciudad?	No creo. No habrá tanto a esta hora.	1
¿Cuántos años tiene Julio?	No sé exactamente. Tendrá unos veinte años.	1
Mamá, ¿puedo jugar con mis amigos?	No, terminarás tus deberes y te quedarás en casa.	1
¡Buenas tardes! ¿Querrían una mesa?	Sí, para dos personas.	1
¿Les parece bien esta?	Muy bien, gracias.	1
¿Desearían comer algo?	Sí.	1
¿Podría traernos la carta?	¡Claro! Un momento, por favor.	1
Aquí la tiene... ¿Qué van a comer?	Para mí, una ensalada de pollo sin cebolla y, de segundo, filete de ternera con patatas fritas.	1
Para mí, una ensalada de pollo sin cebolla y, de segundo, filete de ternera con patatas fritas.	Yo quiero, de primero, un gazpacho y, de segundo, pollo asado con arroz.	1
Yo quiero, de primero, un gazpacho y, de segundo, pollo asado con arroz.	¿Y para beber?	1
¿Y para beber?	Para beber, una botella de agua mineral sin gas.	1
Para beber, una botella de agua mineral sin gas.	Yo, una copa de vino tinto.	1
Yo, una copa de vino tinto.	Muy bien, señoritas, enseguida les traigo.	1
¿Van a tomar postre? Tenemos fruta, flan y helado de chocolate o de vainilla.	Sí, gracias. Para mí, un flan.	1
Sí, gracias. Para mí, un flan.	Para mí, un helado de chocolate.	1
¿Quieren café o té?	Un café solo, por favor.	1
Un café solo, por favor.	Un café con leche, para mí.	1
¿Desean algo más?	No, nada más. Eso es todo. ¡Gracias! La cuenta, por favor.	1
¿Qué te dijo María?	Me dijo que vendría el lunes.	1
¿Por qué no estabas en casa?	Porque pensaba que volverías más tarde.	1
¿Qué hora sería cuando salió Juan?	Serían las once de la noche.	1
¿Con quién hablaría Ana ayer?	Hablaría con Mario.	1

Table A13: Examples of labeled data: Textbook (13)

¿Podrías ayudarme?	¡Por supuesto!	1
¿Qué les apetecería comer?	Nos gustaría pedir un plato combinado.	1
¿Te gustaría ir a ver una exposición de Goya?	Sí, me encantaría.	1
¿Te importaría abrir la ventana?	¡Claro que no!	1
Voy a comprar una moto.	Yo que tú, no lo haría. Compraría un coche.	1
Disculpe, ¿cómo se puede ir al Museo del Prado desde aquí?	¡Es fácil! Puede ir caminando. Está muy cerca.	1
¡Es fácil! Puede ir caminando. Está muy cerca.	¡Qué bueno!	1
¿Podría explicarme cómo llegar?	¡Sí, cómo no! Primero diríjase hacia el Paseo de Recoletos, luego camine por el Paseo de Recoletos hasta llegar a la Plaza de Cibeles.	1
¡Sí, cómo no! Primero diríjase hacia el Paseo de Recoletos, luego camine por el Paseo de Recoletos hasta llegar a la Plaza de Cibeles.	¡Perfecto! ¿Y luego?	1
¡Perfecto! ¿Y luego?	Coja a la izquierda hacia el Paseo del Prado.	1
Coja a la izquierda hacia el Paseo del Prado.	Bien, bien, entiendo...	1
Siga caminando por el Paseo del Prado, unas cuatro calles, luego gire a la izquierda hacia la Calle Felipe IV. El Museo del Prado está exactamente a la derecha de esta calle.	¡Muchas gracias!	1
Con este mapa y las indicaciones que usted me ha dado no habrá ningún problema en llegar.	¡Seguro que no!	1
¿Puedo pasar?	Sí, pasa, pasa.	1
¿Podemos abrir la ventana?	Sí, abrid la ventana, por favor.	1
¿Puedo pasar?	Sí, pase, por favor.	1
¿Podemos cerrar la puerta?	Sí, cierren la puerta, por favor.	1
¿Puedo hablar con la secretaria?	Sí, habla con ella.	1
¿Puedo hablar con la secretaria?	No, no hables con ella.	1
¿Puedo hablar con la secretaria?	Sí, hable con ella.	1
¿Puedo hablar con la secretaria?	No, no hable con ella, por favor.	1
¿Podemos abrir la ventana?	Sí, abrid la ventana.	1
¿Podemos abrir la ventana?	No, no abráis la ventana.	1
¿Podemos abrir la ventana?	Sí, abran la ventana.	1
¿Podemos abrir la ventana?	No, no abran la ventana, por favor.	1
¿Puedo abrir la ventana?	Sí, ábrela.	1
¿Puedo abrir la ventana?	Sí, ábrala.	1
¿Puedo abrir la ventana?	No, no la abras.	1
¿Puedo abrir la ventana?	No, no la abra.	1
¿Puedo dar el documento a la secretaria?	Sí, dáselo.	1
¿Puedo dar el documento a la secretaria?	Sí, déselo.	1
¿Puedo dar el documento a la secretaria?	No, no se lo des.	1
¿Puedo dar el documento a la secretaria?	No, no se lo dé.	1
Lo siento.	No te preocupes.	1
Lo siento.	No se preocupe.	1
¿Me das el disco?	Sí, toma.	1
¿Me puedes dar el disco?	Sí, toma.	1
¿Me trae un vaso de agua?	Sí, espere un momento.	1
¿Me puede traer un vaso de agua?	Sí, espere un momento.	1
¿Me dejas 100 pesos?	No puedo. No tengo dinero.	1

Table A14: Examples of labeled data: Textbook (14)

¿Me puedes dejar 100 pesos?	No puedo. No tengo dinero.	1
¿Perdone! ¿Podría abrir la ventana?	¡Claro! Ahora mismo la abro.	1
¡Oiga! ¿Perdone! ¿Dónde está la Gran Vía, por favor?	Siga todo recto y tome la segunda calle a la izquierda. Luego, cruce la plaza de Medinaceli.	1
Allí está la Gran Vía.	¡Muchas gracias!	1
¡Oye! ¿Perdonal! ¿Sabes dónde está la estación de metro "Sevilla"?	Sí, sigue todo recto y gira en la primera a la derecha. Después s. toma la calle Gaya y sigue hasta la Glorieta Bilbao.	1
Sí, sigue todo recto y gira en la primera a la derecha. Después, toma la calle Gaya y sigue hasta la Glorieta Bilbao.	¡Vale! ¡Gracias!	1
¿Hay un restaurante por aquí?	Sí, hay uno al final de esta calle, a la izquierda.	1
Sí, hay uno al final de esta calle, a la izquierda.	¿Está lejos?	1
¿Está lejos?	No, está aquí mismo. A unos cinco minutos andando.	1
No, está aquí mismo. A unos cinco minutos andando.	¡Gracias!	1
¿Qué lengua se habla en Canadá?	Se habla francés e inglés.	1
¿Se puede pagar con tarjeta de crédito?	No, no se puede. Aquí se paga solo en efectivo.	1
¿Sí? ¡Diga!	¡Carlos, soy María! Te llamo desde una cabina de la Plaza de España.	1
¿Pasa algo? ¿Estás bien?	Acaban de robarme el bolso.	1
¿Puedes venir por favor? Estoy muy nerviosa y no sé qué hacer.	Sobre todo, te pido que te calmes.	1
Dime, ¿qué tenías en el bolso?	Mi móvil, mi cartera con dinero, tarjetas de débito y crédito, mi DNI, un monedero pequeño y otras cosas más.	1
Te recomiendo que canceles inmediatamente tus tarjetas. Luego, iremos a una comisaría para denunciar el robo. ¡Ahora mismo voy!	¡Vale!	1
Pero, ¿dónde nos vemos?	Espérame al lado del monumento de Miguel de Cervantes. No te muevas de allí.	1
Carlos, ¡gracias por venir!	Vamos rápido, busquemos por el cammo y miremos en las peleras del parque. Generalmente, los ladrones solo quieren el dinero en efectivo.	1
¡Ojalá encontremos tu DNI!	¡Espero que sea así!	1
Después, hay que ir rápidamente a una comisaría.	De acuerdo.	1
Carlos, ¡gracias por ayudarme!	¡De nada! Para eso están los amigos.	1
¿Qué quieres hacer hoy?	Quiero salir con mis amigos, pero mis padres quieren que yo me quede en casa.	1
Espero verte pronto.	Yo también espero que vengas pronto.	1
Hacia dónde partiremos mañana? Ven y veamos el mapa.	Pues, estaba pensando en Toledo.	1
¿Qué te parece? Salgamos temprano, no creo que tardemos mucho en llegar.	De acuerdo, pero mañana conduce tú. Llevo tres días conduciendo y estoy cansado.	1
De acuerdo, pero mañana conduce tú. Llevo tres días conduciendo y estoy cansado.	No te preocupes. Mañana conduciré yo. Solo espero que hag a buen tiempo...	1
En el periódico he visto que mañana hará sol.	¡Qué bien! No te olvides de la cámara, ayer la dejaste en el armario del hotel.	1
¡Qué bien! No te olvides de la cámara, ayer la dejaste en el armario del hotel.	Descuida. Mejor dámela ahora y la meto en la maleta. ¿Y qué vamos a hacer en Toledo?	1
Mira, hay numerosas opciones para recorrer la ciudad. Hay rutas históricas, artísticas y también existen las rutas de don Quijote donde se pueden ver algunos parajes que aparecen en El Quijote.	¡Qué interesante! Entonces, vayamos a visitar algún molino.	1
¿Qué interesante! Entonces, vayamos a visitar algún molino.	Sí, hay uno en Madridejos y, por lo visto, es uno de los más antiguos.	1
Sí, hay uno en Madridejos y, por lo visto, es uno de los más antiguos.	Genial, siempre he querido visitar uno ... Me encanta el capitulo de los molinos de viento de El Quijote.	1
¡Ah! En cuanto lleguemos al hotel, pregunta qué autopista debemos tomar para ir a Toledo.	De acuerdo, luego se lo preguntaré al recepcionista.	1
Ahora disfrutemos de nuestra última noche en Ávila.	Vale. Reserva algún lugar y salgamos a tomar unas copas.	1
¡Carmen, cuánto tiempo sin verte!	¡Hola Rubén! ¿Cómo te va todo?	1
¡Hola Rubén! ¿Cómo te va todo?	Muy bien, gracias.	1
¿Vendrás a jugar al tenis con nosotros el sábado por la mañana, verdad? Espero que no llueva.	Pues dudo que pueda ir ... Estoy haciendo un voluntariado durante los fines de semana.	1
Pues dudo que pueda ir ... Estoy haciendo un voluntariado durante los fines de semana.	¿Ah, sí? No lo sabía. ¿Qué tipo de voluntariado?	1
¿Ah, sí? No lo sabía. ¿Qué tipo de voluntariado?	Ayudo a la gente mayor que vive en una residencia de la tercera edad.	1
Pues, estando solo con ancianos, no creo que el voluntariado sea divertido ...	Al contrario, me encanta ir. Disfruto mucho conversando con ellos.	1

Table A15: Examples of labeled data: Textbook (15)

¿Y de qué hablas con ellos que pueda ser interesante?	Me cuentan historias de cuando eran jóvenes, me hablan de sus hijos y nietos. Es muy entretenido, se me pasa el tiempo volando.	1
Me cuentan historias de cuando eran jóvenes, me hablan de sus hijos y nietos. Es muy entretenido, se me pasa el tiempo volando.	¿Qué interesante!	1
¿Permiten que personas que no sean voluntarias hagan visitas?	Claro que sí. Te aconsejo que vengas la semana que viene. Se celebra una fiesta para recaudar fondos para mejorar los jardines de la residencia.	1
Claro que sí. Te aconsejo que vengas la semana que viene. Se celebra una fiesta para recaudar fondos para mejorar los jardines de la residencia.	Está bien, iré. Yo también estaba pensando en hacer un voluntariado.	1
Está bien, iré. Yo también estaba pensando en hacer un voluntariado.	Genial, dile a Carlos que venga también. Lo pasaremos bien.	1
El filósofo José Ortega y Gasset explicaba que era impensable estudiar la historia de España sin considerar las corridas de toros, ¿no?	Así es, tenemos suerte de que en su libro La caza y los toros se hayan reunido sus escritos sobre el tema.	1
En cuántas fiestas has toreado hasta el momento?	Llevo 50 y presento, é mi alternativa, lo que sería la graduación para los toreros, el próximo domingo en la "Monumental Plaza de Toros México".	1
Rafael, antes de que empecemos, me gustaría que nos explicaras los elementos principales de esta fiesta, para que podamos entenderla mejor.	Con mucho gusto, Pilar. Una corrida de toros es un espectáculo muy diferente a todos los demás y, por otro lado, es también un arte.	1
Con mucho gusto, Pilar. Una corrida de toros es un espectáculo muy diferente a todos los demás y, por otro lado, es también un arte.	Espectáculo y arte, ¿qué interesante!	1
Pero tengo una sorpresa para nuestros amigos televidentes.	¿Qué sorpresa? No me habías comentado nada sobre la empresa...	1
¡Bienvenido, Leandro de Andalucía! Es un honor tenerte en nuestro programa.	Muchas gracias, Pilar.	1
Nuestro amigo Ricardo Velasco pregunta cuál es el momento más difícil de la corrida.	El momento más difícil es durante el último tercio de la corrida, el tercio de la muerte en donde se realiza el mayor enfrentamiento del matador con el toro.	1
He sabido que cortaste una oreja en tu última corrida.	Así es, cuando la corrida le ha gustado al público, la gente comienza a mover simbólicamente sus pañuelos solicitando al presidente del jurado un trofeo para el matador.	1
Qué triste que hayan convertido la antigua plaza de toros de Las Arenas en un centro comercial.	Pues a mí me parece que han hecho un buen trabajo.	1
No puedo creer que hayan prohibido las corridas de toros en Cataluña.	Pues yo me alegro de que el gobierno lo haya decidido así.	1
Se trata de una tradición muy antigua, no sé qué tiene de malo.	Alguien tiene que proteger los derechos de los animales, ¿no crees?	1
Alguien tiene que proteger los derechos de los animales, ¿no crees?	Es una pena que de un tiempo acá la gente haya visto las corridas como actos de violencia, porque más bien se trata de un arte, de un ritual.	1
No entiendo cómo pueden considerarlas manifestaciones artísticas cuando son una práctica de crueldad contra los toros.	Es una lástima que la gente haya interpretado así esta tradición, pero bueno no vamos a discutir más sobre esto.	1
Es una lástima que la gente haya interpretado así esta tradición, pero bueno no vamos a discutir más sobre esto.	Sí, vamos a dejarlo aquí. Me parece que no vamos a ponernos de acuerdo en este tema.	1
Sí, vamos a dejarlo aquí. Me parece que no vamos a ponernos de acuerdo en este tema.	Bueno, ya que estamos aquí, aprovechemos las tiendas que no han cerrado para hacer algunas compras.	1
Bueno, ya que estamos aquí, aprovechemos las tiendas que no han cerrado para hacer algunas compras.	Es una buena idea.	1
Daniel, ¿cómo te fue en la entrevista de ayer para hacer tus prácticas?	¿Cómo sabes que tuve una entrevista?	1
¿Cómo sabes que tuve una entrevista?	Noelia me dijo que te vio por los pasillos de la empresa.	1
Noelia me dijo que te vio por los pasillos de la empresa.	Ah, sí... Me alegré de verla porque cuando llegué no había nadie que me conociera.	1
Me dijo que buscaban a alguien que supiera hablar muchos idiomas.	Sí. Ese es el problema, solo hablo inglés y un poco de alemán.	1
No te desanimes. Aún no conoces los resultados.	Ya, pero si pudiera cambiar mi mala suerte...	1
Ya, pero si pudiera cambiar mi mala suerte...	Si confiaras más en ti mismo, las cosas te irían mejor.	1
Además, uno de los entrevistadores era un antiguo profesor que tuve en el primer año. ¿Y sabes qué? Me miró como si no me conociera.	Hombre, siendo entrevistador es lógico que tuviera que ser objetivo.	1
Sí, pero luego lo volví a ver en el pasillo y me ignoró, pasó como si yo no existiera.	Tal vez no se acuerde de ti.	1
Tal vez no se acuerde de ti.	Puede ser...	1
mejor cambiemos de tema. Voy a dar una fiesta en mi piso, ¿vendrás, no?	¿Cuándo es?	1
¿Cuándo es?	El sábado por la noche.	1
Desearía que vinieras con tu hermana.	¿Mi hermana? ¿Por qué?	1
Daniel, ¿cómo te fue en la entrevista de ayer para hacer tus prácticas?	¿Cómo sabes que tuve una entrevista?	1
¿Cómo sabes que tuve una entrevista?	Noelia me dijo que te vio por los pasillos de la empresa.	1
Noelia me dijo que te vio por los pasillos de la empresa.	Ah, sí... Me alegré de verla porque cuando llegué no había nadie que me conociera.	1
Me dijo que buscaban a alguien que supiera hablar muchos idiomas.	Sí. Ese es el problema, solo hablo inglés y un poco de alemán.	1
No te desanimes. Aún no conoces los resultados.	Ya, pero si pudiera cambiar mi mala suerte...	1
Ya, pero si pudiera cambiar mi mala suerte...	Si confiaras más en ti mismo, las cosas te irían mejor.	1
Además, uno de los entrevistadores era un antiguo profesor que tuve en el primer año. ¿Y sabes qué? Me miró como si no me conociera.	Hombre, siendo entrevistador es lógico que tuviera que ser objetivo.	1

Table A16: Examples of labeled data: Textbook (16)

Sí, pero luego lo volví a ver en el pasillo y me ignoró, pasó como si yo no existiera.	Tal vez no se acuerde de ti.	1
Tal vez no se acuerde de ti.	Puede ser...	1
mejor cambiemos de tema. Voy a dar una fiesta en mi piso. ¿vendrás, no?	¿Cuándo es?	1
¿Cuándo es?	El sábado por la noche.	1
Desearía que vinieras con tu hermana.	¿Mi hermana? ¿Por qué?	1
Ayer, mientras paseaba al perro me encontré a Alicia, nuestra compañera de la secundaria.	¿Ah, sí?	1
¿Y qué hay de su vida? Hace años que no sé nada de ella.	La verdad es que está muy cambiada.	1
¿Ha cambiado mucho?	Ha perdido mucho peso, yo diría que al menos unos diez kilos, está más guapa.	1
Tranquila, me pidió tu número. Te llamará un día de estos.	Perfecto.	1
¿Qué son los perros de asistencia?	Son perros especialmente adiestrados para ayudar a personas con algún tipo de discapacidad o que hubieran sufrido un accidente.	1
No sabía que Alicia hubiera hecho un curso de adiestramiento para perros.	No hace mucho que lo hizo, aunque dijo que le hubiera gustado estudiar veterinaria. Por lo visto, desde siempre le han gustado los animales.	1
¡Hola Cristinal! ¿Cómo te fue en el verano?	Muy bien. He estado en mi pueblo durante todas las vacaciones y disfruté como si hubiera vivido siempre allí.	1
¿No estabas viviendo en un piso compartido cerca de aquí?	Sí, pero ojalá que pudiera dejar el piso.	1
¿Por qué? Decías que era muy económico y bien ubicado.	Sí, pero con tanta gente es imposible estudiar en casa y muchas veces no tengo tiempo para cocinar ni para hacer las tareas del hogar.	1
Entiendo. Pero a estas alturas puede que sea difícil encontrar una habitación disponible.	Lo sé, ojalá tenga suerte y pueda entrar a vivir en la residencia.	1
Hablas como si se tratara de la lotería ...	Pues sí, es algo parecido.	1
Que te asignen una habitación de la residencia es como si te tocara el premio "Gordo" de la lotería.	Bueno, no es de extrañar...	1
En ningún sitio vas a encontrar un precio más bajo y además no creo que ningún restaurante fuera de la residencia ofrezca precios tan baratos.	Ya ... ojalá que alguien cancele su solicitud.	1
Ya ... ojalá que alguien cancele su solicitud.	¡Mucha suerte!	1
¿Qué lees con tanta atención?	Las noticias de hoy.	1
Encontré una noticia en la página de tecnología que me interesa.	¿De qué se trata?	1
Habla de la creciente inquietud en la red por los continuos problemas en el servicio de mensajería móvil gratuita.	Mensajes gratuitos, ¡una maravilla!	1
A ver ... ¿de dónde es la noticia?	Es local.	1
Estoy de acuerdo, pero a veces dependemos demasiado del móvil. ¿no crees?	Tal vez.	1
pero ¿quién podría vivir hoy en día sin un aparato de estos?	Pues, no sé...	1
¿qué estás haciendo aquí? ¿No te vas a casa?	prefiero quedarme aquí.	1
¿Y qué escribes?	Pues estoy escribiendo una carta para la familia que conocí en Bolivia.	1
¿qué país te gustó más?	Viajando por Bolivia, tuve la suerte de hacer muchos amigos. Acabé enamorándome de sus paisajes, pero sobre todo de su gente.	1
El otro día estaba leyendo que no era un país muy visitado por extranjeros.	Es cierto.	1
¿Y cuál fue tu paisaje favorito? Conociéndote, ¡debes haber tomado cientos de fotos!	Me encantaron las ruinas de Tiwanaku y el Salar de Uyuni, pero el Lago Titicaca me impresionó mucho.	1
¿Cómo estás? No tienes buena cara.	Estoy cansada, casi muerta ...	1
¿Qué hiciste el fin de semana?	Fuimos a una exposición sobre culturas prehispánicas.	1
¿Te gustó la exposición?	Sí, ya sabes que este tipo de eventos me llama mucho la atención.	1
Había piezas y obras de arte muy interesantes, también había varias momias encontradas en Los Andes y otros restos hallados en yacimientos de Mesoamérica.	¿Desde dónde trajeron las piezas?	1
¿Desde dónde trajeron las piezas?	Todas las piezas fueron prestadas tanto por museos como por colecciones privadas de Guatemala, Perú y México.	1
Todas las piezas fueron prestadas tanto por museos como por colecciones privadas de Guatemala, Perú y México.	Si tengo tiempo el próximo fin de semana, iré con Carlos.	1
¿No viene Miguel?	No, he venido sola.	1
Necesitaba estar sola para pensar un poco.	¿Qué te pasa? ¿Tenéis algún problema?	1
¿Qué te pasa? ¿Tenéis algún problema?	Tengo algo que contarte.	1
¿Habéis discutido?	No, y creo que ese es el problema.	1

Table A17: Examples of labeled data: Textbook (17)

Es importante hablar cara a cara para solucionar los problemas.	Para ser te sincera, ya no confío en él.	1
¿Qué quieres decir? ¿Que te engaña?	No, bueno ... no lo sé.	1
Perdón por molestarte con mis problemas.	Ni hablar. No te preocupes.	1
¿Dónde va a tener lugar la celebración?	En un hotel que está a las afueras de la ciudad.	1
¿Y qué parte de la celebración te gusta más?	A mí me encanta la parte en la que la quinceañera se cambia de zapatos.	1
Espero que se lo pasen de maravilla.	Muchas gracias.	1
¿Te acuerdas de Ana, la chica cuyo padre tenía una zapatería cerca de la plaza?	Claro que sí.	1
Esa chica de quien me hablas fue mi primer amor.	¡No me digas! No lo sabía.	1
Oye, la próxima vez tenemos que ir juntos.	Y de ser posible, durante las fiestas del pueblo ...	1
Y de ser posible, durante las fiestas del pueblo ...	Buena idea.	1
¿Por qué no avisamos a los demás?	Perfecto.	1
A ver si podemos reunirnos todos, como en los viejos tiempos.	Estaría bien que los que tuviéramos tiempo y quisiéramos ir, alquiláramos una cabaña para pasar las fiestas juntos.	1
Necesito información.	Esa lección fue explicada por el profesor la semana pasada.	1
Esa lección fue explicada por el profesor la semana pasada.	Ya, pero no asistí a clase.	1
¿Puedes echarme una mano?	¡Claro!	1
Pero, ¿tenías el ordenador estropeado, no es verdad?	Sí, pero el problema fue detectado por un técnico y ya está solucionado.	1
¿Sabes dónde se encuentran localizados?	El área geográfica de influencia de estos pueblos indígenas está distribuida en todo el país, específicamente por el occidente, el sur y la zona central.	1
¿Tienes datos sobre la población?	Los datos proporcionados fueron considerados poco confiables.	1
Oye, en vez de estar aquí en casa sin hacer nada, mejor salimos a cenar.	Sí, me estoy durmiendo.	1
Si quieres vamos a ese restaurante chileno que te gusta tanto.	¡Genial! Es el restaurante que me gusta más que ningún otro.	1
Podemos avisar a Rosa y a Luis para ir juntos.	Vale, me parece bien.	1
¿Sabías que en una de las salas venden piezas de artesanía de Chile?	¿Qué tipo de artesanía?	1
¿Qué tipo de artesanía?	Sobre todo piezas realizadas con lapislázuli.	1
Sobre todo piezas realizadas con lapislázuli.	¿Qué es eso?	1
¿Cuándo fundaron ese restaurante? Parece antiguo ...	Pues no lo es. Que yo sepa, no tiene más de diez años.	1
Oye, antes de seguir hablando de otras cosas, más vale que llamemos para reservar una mesa.	Tienes razón	1
¿En qué puedo ayudarles?	Buenos días, quisiéramos información sobre las visitas programadas a la mezquita.	1
Buenos días, quisiéramos información sobre las visitas programadas a la mezquita.	Muy bien, aquí tienen un folleto informativo.	1
¿Cuánto tiempo duraría la visita y en cuánto nos saldría?	La tarifa es de setenta euros y ustedes pueden estar todo el tiempo que quieran sin otro límite que el horario de la mezquita.	1
¿Y cómo son las visitas en grupo?	El precio es de quince euros por persona y la visita dura alrededor de tres horas.	1
Pero, ¿podemos pagar con la tarjeta de crédito, no?	La he dejado en el hotel. Además, setenta euros me parece excesivo.	1
La he dejado en el hotel. Además, setenta euros me parece excesivo.	Es un poco caro, pero ya sabes que tenía muchas ganas de ver esta mezquita.	1
Vaya, no sabía que te estabas aburriendo tanto ...	No te había dicho nada hasta ahora porque sé que te encanta el arte.	1
No te había dicho nada hasta ahora porque sé que te encanta el arte.	Bueno, a partir de mañana cambiaremos los planes...	1
¿Dónde está Luis?	Está en camino.	1
¿Me trajiste el libro que te pedí?	¡Ah! Se lo presté a Luis la semana pasada.	1
Bueno, mejor tarde que nunca. ¿no?	Sí. ... Bueno, ¿y tú qué?	1
¿Ya has decidido qué vas a hacer?	Pues la verdad es que estoy hecho un lío.	1
Pero, ¿no estabas pensando en buscar un trabajo?	He estado mirando y no hay muchas ofertas en lo mío...	1
¿Y cómo te pagarías los gastos si no trabajas?	Tengo algo de dinero ahorrado...	1

Table A18: Examples of labeled data: Textbook (18)

¡Quítate de la puerta!	¡Pasa si puedes! ¡No te dejaré salir!	1
¡Pasa si puedes! ¡No te dejaré salir!	¡Apártate y déjame pasar!	1
¿Pero lo habéis matado?	¡No! ¡Salió corriendo como un cobarde en su caballo!	1
¿Por qué lo has dicho entonces?	¡Por ella!	1
¡Abre, Adela! ¡Abre la puerta!	Abre, Adela.	1
¡No entres!	No. ¡Yo no!	1
¡Lorena! Menos mal que estás aquí. ¡Qué alegría encontrarte!	¡Qué sorpresa! ¡Qué haces aquí!	1
¿No habíamos quedado mañana por la noche?	Sí, sí. Pero no podía esperar...	1
Tengo que hablar contigo urgentemente.	¿Qué te pasa? Te noto un poco nerviosa.	1
¿No puedo seguir adelante con la boda!	¿Cómo que no puedes seguir con la boda! Pero, ¡qué dices! A ver, cálmate.	1
Son los nervios ... , últimamente has tenido mucho trabajo y con todos los preparativos es normal que estés estresada.	No es eso ...	1
no estoy segura de querer pasar el resto de mi vida con Pablo.	¡Es increíble!	1
La boda se celebra dentro de un mes ...	¡Ya lo sé!	1
Entonces, ¿cuál es el problema?	Pues, no sé ... , empiezo a tener dudas sobre mis sentimientos hacia Pablo.	1
Pues, no sé ... , empiezo a tener dudas sobre mis sentimientos hacia Pablo.	¿Has hablado con él?	1
¿Has hablado con él?	¿Qué va! ¿Estás loca?	1
Él está muy ilusionado con la boda y no quiero hacerle daño.	Lo entiendo, pero más vale que aclares tus ideas lo antes posible.	1
Habla con él hoy mismo.	Y ¿qué le digo?	1
Y ¿qué le digo?	Lo mismo que me has dicho a mí.	1
Si necesitas más tiempo, podrían aplazar la boda.	¿Qué fácil es decirlo! Además, ¡casi todo está listo!	1
¡Me voy a volver loca! Me siento tan mal...	Ya, lo sé.	1
Pero casarte con él sin estar convencida no es la solución.	Bueno, luego te llamo.	1
Gracias por todo.	¡Para qué son las amigas!	1
Acabo de cruzarme con un gato negro, espero que no me traiga mala suerte.	No sabía que fueras supersticiosa.	1
No sabía que fueras supersticiosa.	Bueno, no lo soy. Pero mi madre siempre decía: "¡Más vale no tentar a la suerte!".	1
Y hoy la necesitas más que nunca, ¿no?	¡Sí!	1
Seguro que apruebas. Después de todo lo que has estudiado, no tendrás ningún problema.	Eso espero.	1
Están deseosos de que me licencie y que busque un trabajo lo antes posible.	Es normal.	1
Tus hermanos y tú están estudiando. Debe ser una carga muy grande para tus padres.	Es cierto.	1
Cambiando de tema, ¿has visto a Víctor en estos días?	Ayer, precisamente, quedé con él para tomar algo. Me dijo que estaba pensando en buscar otro trabajo.	1
¿Por qué? Lleva menos de un año en esa empresa.	Sí, pero me comentó que se llevaba muy mal con su jefe y que no aguantaba más.	1
¿Qué te pasa? Pareces de mal humor, ¿no te ha gustado la película?	¿Qué va! Si me ha encantado, ya sabes que me interesa muchísimo este tipo de películas.	1
¿Entonces? ¿Por qué estás enfadada?	Pues porque no he podido disfrutar de la película como hubiera querido.	1
Me molesta que la gente deje los móviles encendidos dentro del cine.	Ah, ahora entiendo. Lo dices por el chico que estaba sentado a tu lado. Le ha sonado el teléfono unas cuantas veces.	1
Un par de veces no me importa, pero estuvo enviando mensajes durante toda la película.	Entonces no me extraña que estés malhumorada.	1
Bueno, olvídalos. Te invito a cenar.	Gracias.	1
¿Vamos al italiano de la esquina? Me apetece comer pasta.	Me parece bien, me agrada mucho ese restaurante.	1
Recuerda que a las diez quedamos con Ana y Pepe para ir a tomar unas copas al bar.	¡Lo había olvidado por completo!	1
¿Dónde pongo la botella de plástico?	Pues dondequiera, en la papelería si gustas.	1
¿Con los papeles?	Sí, sí.	1

Table A19: Examples of labeled data: Textbook (19)

María, no me digas que a estas alturas aún no clasificas la basura cuando la tiras.	Es algo que nunca hago.	1
¡Ahora resulta que yo soy la culpable de los problemas del medio ambiente!	No te lo tomes a broma. Es un tema muy grave y alguien tiene que ponerle remedio.	1
Cualquier acción, por pequeña que sea, contribuirá a mejorar este planeta.	Bueno, no sabía que estuvieras a favor del movimiento ecológico.	1
Buenas tardes, ¿es esta la residencia de la escuela de idiomas?	Sí, aquí es.	1
Me inscribí por Internet. Mi nombre es Carolina Hernández.	No encuentro su inscripción. ¿De dónde viene?, ¿cuál es su segundo apellido?	1
No encuentro su inscripción. ¿De dónde viene?, ¿cuál es su segundo apellido?	Vengo de Paraguay. Mi apellido materno es Choi.	1
Vengo de Paraguay. Mi apellido materno es Choi.	¿Cómo se deletrea?	1
¿Cómo se deletrea?	Che (de chico), o (de Oscar), i (de iglesia).	1
Un momento por favor..., aquí está. Estaba en la lista de pendientes.	Pero yo pagué la inscripción. Aquí está el comprobante.	1
Pero yo pagué la inscripción. Aquí está el comprobante.	No hay problema. Su número de pasaporte, por favor.	1
No hay problema. Su número de pasaporte, por favor.	Cero, Cero, cuarenta y ocho, setenta y cinco, diez	1
Va a estar en la quinientos quince.	¿Es individual o compartida?	1
¿Es individual o compartida?	La que usted pagó era compartida. Su compañera llegó esta mañana.	1
Muchas gracias, muy amable.	De nada. Lo que se ofrezca, con gusto la ayudaremos.	1
¿Qué desea?	Quisiera solicitar una extensión de mi visa. Caduca en quince días.	1
Quisiera solicitar una extensión de mi visa. Caduca en quince días.	Un momento... ¿Apellido?	1
Un momento... ¿Apellido?	Lee.	1
¿Apellido materno?	No aparece en mi pasaporte. Pero es kim.	1
¿Fecha y lugar de nacimiento?	El 16 de mayo de 1983, en Seúl.	1
¿Nacionalidad? ¿Estado civil?	Coreana. Soltera.	1
¿Dónde vive actualmente? Domicilio y teléfono.	Aquí, en Madrid. Calle Gante, número 67, segundo-A. Treinta y ocho, veinticinco, quince, cero cuatro.	1
Aquí siempre preguntan el apellido materno, ¿te fijas?	¿Y qué más te preguntó?	1
¿Y qué más te preguntó?	Me preguntó que dónde y cuándo había nacido y mi estado civil. También me preguntó que dónde vivía.	1
Parecía enfadado.	Yo pensaba igual al verlo, pero fue muy amable y dijo que esperaba a que me llamaran.	1
¿Adónde tenemos que ir primero?	Primero tenemos que pasar por donde revisan los pasaportes.	1
Primero tenemos que pasar por donde revisan los pasaportes.	Aquí está el letrero, es por el pasillo de la izquierda.	1
Es mejor tener el pasaporte listo junto con el formulario que llenamos en el avión, ¿lo tienes ya?	No encuentro el formulario, ¿dónde lo puse? ¡Ah! Aquí estaba. Pensé que lo había puesto en mi cartera.	1
Sus pasaportes, por favor. ¿Son ciudadanas de la Unión Europea?	Sí, señor. Somos españoles.	1
¿Cuál es el objetivo de su viaje?	Venimos de turistas. Estamos de vacaciones.	1
Venimos de turistas. Estamos de vacaciones.	¿Es su primera visita a México? ¿Adónde piensan viajar?	1
¿Es su primera visita a México? ¿Adónde piensan viajar?	Sí, es nuestra primera visita. Queremos conocer muchos lugares de México, especialmente Cancún.	1
Sí, es nuestra primera visita. Queremos conocer muchos lugares de México, especialmente Cancún.	Muy bien. Todo está en orden. Feliz estancia y que disfruten en nuestro país.	1
Quisiera pasar al servicio, ¿en dónde nos vemos?	Donde tú digas.	1
¿Qué te parece en las bandas para recoger el equipaje?	Muy bien. No tardes.	1
¿Cuáles son sus maletas?	Son estas. Y esas son las de mi amiga.	1
Pongan aquí su equipaje. ¿Cuántas maletas traen en total?	Cuatro, dos grandes, dos pequeñas y nuestras mochilas.	1
Denme por favor sus declaraciones. ¿Traen algo en especial?	No, señor. Nada.	1
¿De dónde vienen?	De España.	1
Abran esta maleta grande, por favor.	Sí, como no. Ana, ayúdame ¿sí? Detenme mi mochila.	1
¿Llevan tabaco o licores?	Solo son botellas de sangría, y unos paquetes de cigarros.	1

Table A20: Examples of labeled data: Textbook (20)

A ver, yo le ayudo. Ya está. Pasen, por favor.	Muchas gracias.	1
¿En qué vuelo ha llegado usted?	En el vuelo CA219, proveniente de Pekín.	1
¿Lleva algo de valor en su maleta?	No, señor. Solo ropa y objetos personales.	1
Describame su maleta, por favor.	Es rectangular, de color negro. Tiene ruedas y viene cerrada con un candado pequeño. Le he puesto dos listones para distinguirla, uno rojo y otro azul.	1
Bien. Un momento ... Aquí clave 34. Rectangular, negra con ruedas y candado. Listones azul y rojo. Cambio. Afinnativo. Se encuentra conmi go. Cambio. Esperaremos. Cambio y fuera.	¿Han encontrado mi maleta?	1
¿Han encontrado mi maleta?	Hemos encontrado una que tal vez sea la suya. Aquí viene.	1
¿No me ha dicho usted que no traía nada de valor?	Pensé que se refería a joyas o dinero. Disculpe.	1
No tuve tiempo por la mañana.	A mí me pasó igual.	1
Buenas tardes. ¿A qué hora sale el próximo tren para Barcelona?	Un momento. Hay un AVE que sale a las doce quince.	1
Un momento. Hay un AVE que sale a las doce quince.	Sí, ese está bien.	1
Sí, ese está bien.	Perfecto.	1
Deme dos pasajes para el AVE de las doce quince y un itinerario.	Mejor pídele dos itinerarios, uno para mí también.	1
¿Billetes de ida y vuelta o sencillos?	De ida y vuelta, por favor.	1
¿En qué andén sale este tren?	En el número seis. Quince minutos antes de la salida deben esperar allí.	1
¡Caramba! Tras cobrar tan caros los billetes todavía tuvimos que pagarle un seguro.	No te quejes tanto. En lugar de gastar tus energías en eso, empieza a disfrutar desde ahora.	1
No te quejes tanto. En lugar de gastar tus energías en eso, empieza a disfrutar desde ahora.	Tienes razón.	1
Buenas tardes, señorita. Pase a mi escritorio, por favor. ¿En qué puedo servirle?	Quiero reservar un pasaje de ida y vuelta a Seúl. ¿Cuándo hay vuelos?	1
Quiero reservar un pasaje de ida y vuelta a Seúl. ¿Cuándo hay vuelos?	Salen vuelos todos los días. ¿Desea alguna excursión? ¿Cuándo tiene pensado viajar?	1
Salen vuelos todos los días. ¿Desea alguna excursión? ¿Cuándo tiene pensado viajar?	En agosto y, no, no creo que sea muy buena para mí una excursión. ¿Es un vuelo directo o debo transbordar?	1
En agosto y, no, no creo que sea muy buena para mí una excursión. ¿Es un vuelo directo o debo transbordar?	Hace escala en París, pero no hay que cambiar de avión.	1
Hace escala en París, pero no hay que cambiar de avión.	Bien.	1
¿Cuántas horas dura el viaje?	Aproximadamente, catorce horas.	1
¿Desea un pasaje de primera clase?	No, de turista.	1
¿Con cuánta anticipación debo reservar el pasaje?	En esta semana, si fuera posible. Es temporada alta y es probable que los pasajes de clase turista se acaben pronto.	1
En esta semana, si fuera posible. Es temporada alta y es probable que los pasajes de clase turista se acaben pronto.	Entonces, hágame el favor de reservar uno para el 12 de agosto.	1
¿Hasta cuándo debo confirmar la reservación?	En tres días, máximo.	1
Perdone la pregunta, ¿va usted de paseo a Seúl? ¡Está muy lejos!	No, realmente voy para estudiar coreano.	1
Buenos días. Me permite su boleto, por favor. ¿Prefiere un asiento en ventanilla o en pasillo?	Un asiento en ventanilla, en la sección de no fumar, por favor.	1
¿Va a documentar algún equipaje?	Sí, tengo tres maletas y un maletín.	1
Lo siento, solo puede documentar dos maletas que no excedan los 30 kilos. El exceso de equipaje se cobra extra.	Entendido. Entonces, solo documente las dos maletas grandes. ¿Puedo llevar la pequeña y el maletín en la mano?	1
¿Cuáles son?	La maleta pequeña roja y el maletín negro.	1
¿Señorita, olvida algo!	¿Sí, dígame?	1
Aquí hay un comprobante ...	No es mío, los comprobantes de mi equipaje están aquí.	1
¡Hola! ¡Felicidades!	Gracias. Paco. Pasa, por favor.	1
¡Feliz cumpleaños! Tengo un regalo para ti.	Hummm ... Muchas gracias. No hacía falta regalarme algo. A ver, a ver, ¿qué es? ¡Un oso de peluche!	1
Hummm ... Muchas gracias. No hacía falta regalarme algo. A ver, a ver, ¿qué es? ¡Un oso de peluche!	¿Te gusta?	1
¿Soy el primero en llegar? ¿Dónde están todos?	Están acá. Mira, te voy a presentar a unos amigos. Ella es mi prima Bety.	1
Están acá. Mira, te voy a presentar a unos amigos. Ella es mi prima Bety.	Mucho gusto. Francisco Ramírez.	1
Mucho gusto. Francisco Ramírez.	Bertha Suárez. Mucho gusto.	1
¡Ah! Allá está Cristina, es compañera de Alicia y también es amiga mía.	¿La rubia?	1

Table A21: Examples of labeled data: Textbook (21)

Si, claro. Ven, Paco, vamos por nuestra copa.	Vamos a brindar ... ¡Salud!	1
Vamos a brindar ... ¡Salud!	¡Salud!	1
Joven, ¿sabe usted cuánto tiempo dura el vuelo?	Dura unas quince horas.	1
¿Van a pasar alguna película?	Van a pasar dos, una después de la comida y otra después de la cena.	1
¿Qué desea para comer? Tenemos filete de ternera y pescado con arroz.	Pescado, por favor.	1
¿Qué le sirvo de tomar?	Vino blanco, por favor.	1
Vino blanco, por favor.	Aquí tiene.	1
¿Necesita algo más?	Nada más. Gracias.	1
¡Ah, sí, un momento! ¿Me puede dar una pastilla para el mareo? Siempre me siento un poco mareada cuando hay turbulencia.	Sí, claro. Enseguida se la traigo.	1
Aquí tiene, señora.	Muchas gracias.	1
Oiga, joven, ¿me podría ayudar a llenar los formularios de migración? Yo no sé francés ...	Con mucho gusto. Vendré para ayudarlo cuando distribuya los formularios.	1
Buenas tardes, a sus órdenes.	Buenas tardes. Desearía una habitación para dos personas para este fin de semana.	1
¿Tiene habitaciones libres?	Un momentito, por favor ... ¿Querría usted una habitación doble?	1
Un momentito, por favor ... ¿Querría usted una habitación doble?	Sí, así es, con vista al mar si es posible.	1
¿Prefiere una cama matrimonial o dos individuales?	Preferimos una cama matrimonial.	1
¿Cuánto cobra por noche?	135 euros por noche y 25 euros extra por cada persona adicional.	1
Serían 145 euros por noche.	¿Pensión completa incluida?	1
¿Pensión completa incluida?	No, señor. Este precio solo incluye el desayuno, pero no considera los servicios ni los impuestos.	1
¿Acepta tarjetas de crédito y cheques de viajero?	Sí, señor.	1
¿Podría enviar una botella de champán a la habitación?	Con mucho gusto.	1
Podríamos enviar al jefe de camareros para que ordene su cena.	No es necesario que vaya a nuestra habitación. Bajaremos a la restauración.	1
¿Quién toca?	Buenos días. Venía a ver el piso. ¿Podría verlo ahora?	1
Buenos días. Venía a ver el piso. ¿Podría verlo ahora?	Sí, adelante, por favor.	1
¿Cuántos dormitorios hay?	Dos. Aquí, al lado, están el salón y la cocina.	1
Podrá ver que está totalmente amueblado.	¡Menos mal! Yo no tengo muchos muebles.	1
Y, ¿podría ver el cuarto de baño?	Sí, por supuesto. Está allí, por el pasillo a la izquierda. Hay agua caliente y fría. También hay una ducha y una bañera grande para hidromasaje.	1
Sí, por supuesto. Está allí, por el pasillo a la izquierda. Hay agua caliente y fría. También hay una ducha y una bañera grande para hidromasaje.	¿Qué bien! ¿Hay calefacción? En invierno siempre baja la temperatura por las noches.	1
¿Qué bien! ¿Hay calefacción? En invierno siempre baja la temperatura por las noches.	Naturalmente.	1
¿Cuánto cuesta el alquiler?	En principio, deberá hacer un depósito de 500 euros, y deberá pagar por adelantado dos mensualidades. Cada mes tendrá que pagar 450 euros, gastos aparte.	1
En principio, deberá hacer un depósito de 500 euros, y deberá pagar por adelantado dos mensualidades. Cada mes tendrá que pagar 450 euros, gastos aparte.	Entendido. Entonces, vendré mañana por la mañana para hacer el depósito a su cuenta.	1
Entendido. Entonces, vendré mañana por la mañana para hacer el depósito a su cuenta.	Muy bien. Llamaré a mi abogado para que prepare el contrato.	1
Quisiéramos rentar un auto.	Sí, señor. ¿Qué tipo de auto desearía?	1
Sí, señor. ¿Qué tipo de auto desearía?	Me gustaría un cano compacto de cuatro puertas.	1
Me gustaría un cano compacto de cuatro puertas.	Como modelo compacto de cuatro puertas tenemos el Opel Corsa y el Century Clío. Los dos tienen cuatro asientos y aire acondicionado.	1
¿Son automáticos?	Los tenemos con transmisión automática y también manual.	1
¿Cuál es la tarifa?, para ver cuál nos conviene ... ¿Cobran por kilometraje?	Depende. Si lo alquila por un día, sí; cobramos 5 euros por cada 10 kilómetros.	1
Depende. Si lo alquila por un día, sí; cobramos 5 euros por cada 10 kilómetros.	Muy bien.	1
¿Le incluyo el seguro en la factura?	Sí, claro, incluya el seguro contra todo riesgo.	1
Pennitame su licencia de conducir.	Nosotros somos ciudadanos mexicanos. ¿Necesitamos un permiso especial para conducir en España?	1
Nosotros somos ciudadanos mexicanos. ¿Necesitamos un permiso especial para conducir en España?	No, señor. Su licencia local para conducir es suficiente.	1

Table A22: Examples of labeled data: Textbook (22)

Disculpe, señorita, aquí en España cuando se saludan se dan dos besos. ¿verdad?	Así, es. Es una costumbre muy general entre los amigos y familiares. ¿En México, no?	1
Así, es. Es una costumbre muy general entre los amigos y familiares. ¿En México, no?	Sí, pero solo nos saludamos de mano o nos damos un beso, nada más.	1
¿Diga?	Hola, buenos días. ¿Podría hablar con Andrea, por favor?	1
Hola, buenos días. ¿Podría hablar con Andrea, por favor?	¿Con quién? Creo que usted se ha equivocado. Aquí no vive ninguna Andrea.	1
¿Con quién? Creo que usted se ha equivocado. Aquí no vive ninguna Andrea.	Ay, he marcado otro número, disculpe.	1
¿Sí, diga?	Perdone, ¿podría hablar con Andrea, por favor?	1
Perdone, ¿podría hablar con Andrea, por favor?	En este momento no está. ¿De parte de quién?	1
En este momento no está. ¿De parte de quién?	De Estela, su compañera de trabajo.	1
¿Es algo urgente? ¿Desea dejar algún recado?	Sí, por favor. Se trata de una reunión de la facultad. Necesito avisarle de un cambio antes de que haga otro compromiso.	1
Por favor, le dice que en vez de ser la reunión la semana próxima, será	mañana, miércoles, a las 6 de la tarde.	1
mañana, miércoles, a las 6 de la tarde.	... mañana, miércoles, 6 de la tarde, facultad ...	1
¿Algo más?	Nada más, gracias.	1
De todas formas, volveré a llamarle para confirmar.	Como quieras. Andrea regresará más o menos a la hora de comer.	1
Tan pronto como vuelva, le daré tu mensaje.	Muy bien, señor.	1
Disculpe, ¿con quién tengo el gusto de hablar?	Con el tío de Andrea, soy su tío Luis.	1
Con el tío de Andrea, soy su tío Luis.	Muchas gracias, don Luis.	1
¿Sabías que hemos quedado los de la clase para cenar este miércoles por la noche?	¿Sí? Pues, no, no lo sabía.	1
¿Quisieras venir?	No tengo nada que hacer mañana... Me encantaría ir.	1
Está timbrando pero... no contesta.	Ayer cuando la llamé estaba practicando aeróbicos, y era a esta hora.	1
Ayer cuando la llamé estaba practicando aeróbicos, y era a esta hora.	Sí, tal vez por eso no contesta.	1
¿Te has comprado un celular nuevo? Parece un modelo muy moderno.	Sí, lo es. Estaba en oferta cuando lo compré.	1
El que tenías era muy bueno también. ¿no?	Sí, pero me daba un poco de problemas la batería.	1
Información, buenas tardes.	Buenas tardes.	1
Me han robado la maleta. Por eso, necesito el número de teléfono de la embajada de Corea y el número de la tarjeta VISA.	Un momento, por favor ...	1
¿Cómo puedo hacer una llamada a Corea? He intentado pero no pude comunicarme.	Marque el 07 y después 82, enseguida marque el código de la ciudad.	1
¿Y para llamar a cobro revertido?	Dígame el número y su nombre, por favor.	1
¿Con quién desea hablar?	Con quien conteste.	1
Muy bien, cuelgue y espere un momento.	De acuerdo. Gracias.	1
¿Le puedo ayudar? ¿Qué desea?	Estoy viendo los vestidos.	1
¿Qué modelo necesita? ¿Cómo lo prefiere?	Pues ... uno de color luminoso y alegre, pero un poco menos brillante que ese azul, y que también sea de seda.	1
¿Qué talla le doy?	La treinta y seis.	1
Mmmm ... Mire, tenemos estos modelos. ¿Le gusta alguno?	Este. Me gusta este.	1
¿Puedo probármelo?	Por supuesto. El probador está al fondo, a la izquierda.	1
¿Qué tal le queda?	Me queda un poco apretado. ¿No tiene otro igual en una talla más grande que esta?	1
Me queda un poco apretado. ¿No tiene otro igual en una talla más grande que esta?	No, lo siento.	1
¿Y el precio? Me dijo que era más caro que el otro.	No tanto.	1
Costaba 120 euros, pero como ahora estamos en rebajas, cuesta solamente 79 euros. Es una ganga.	Bien. Me lo llevo.	1
Necesito ahora unos zapatos tan elegantes como el vestido.	Muy bien.	1
¿En qué número?	Del número 24.	1
¿Qué le parecen estos?	Me quedan bien. Me gustan ... Me los llevo.	1

Table A23: Examples of labeled data: Textbook (23)

Me quedan bien. Me gustan ... Me los llevo.	¿Cómo va a pagar?	1
¿Cómo va a pagar?	Con tarjeta.	1
Buenos días, ¿cómo le va, señorita Silvia?	Buenos días, don Pepe.	1
¿Cuánta gente tiene hoy! Vuelvo más tarde.	No, solo es cuestión de unos minutos y ahora estoy con usted.	1
No se vaya.	Bueno, entonces, espero un ratito.	1
Mientras, voy a ir escogiendo algo de aquí.	Bien, ya estoy con usted.	1
¿Qué se va a llevar hoy?	No tengo nada en la nevera, acabo de regresar de vacaciones, así que necesito de todo.	1
¿A cómo están las patatas?	A tres euros el kilo.	1
¿Qué frutas están buenas hoy?	Todas, toda la fruta está muy buena, señorita Silvia. Pero le recomiendo las cerezas. Están bien maduras.	1
Pruebe una.	Hum ... Sí, está riquísima.	1
¿A cuánto está el kilo?	A 5 euros.	1
¿Algo más?	Nada más.	1
Aquí tiene, 35 euros. Y los 10 de la cuenta pendiente de la otra vez.	Muchas gracias.	1
Que pase un muy buen día, señorita Silvia.	Igualmente, don Pepe. Gracias.	1
Restaurante La Vianda, buenos días.	Buenos días.	1
Disculpe, he reservado una mesa hace un momento a nombre de la señora Gil y quería hacer un cambio.	Sí, aquí está. A las 9, para dos personas.	1
¿No hay problema?	No hay ningún inconveniente.	1
Buenas noches, señores. ¿Tienen reservación?	Sí, a nombre de la señora Gil.	1
Nos gustaría una mesa al lado de la ventana.	Muy bien. Por aquí, por favor.	1
¿Qué lugar tan exótico! ¿No será demasiado caro?	No, no es tan caro. Espero que les guste.	1
Aquí tienen la carta y la lista de vinos. ¿Les apetece un aperitivo?	Sí, tomaremos vermut y jamón serrano.	1
Yo no bebo. Para mí, una sangría, si no fuera mucha molestia.	Una sangría para la señora.	1
¿Y para beber?	Agua mineral con gas y tres copas de vino tinto.	1
Camarero, la cuenta, por favor.	Aquí la tiene.	1
¿Ha sido de su gusto la cena?	¡Exquisita! ...	1
Perdone. Creo que no está bien la cuenta. No hemos tomado postre.	Les ruego que nos disculpen. Ha sido un error de facturación.	1
¡Hola, Beto! ¡Cuánto me alegro de verte! Oye, ¿no tienes algún compromiso ahora?	No, solo estaba dando una vuelta por aquí.	1
No, solo estaba dando una vuelta por aquí.	Pues, entonces, ¿tienes tiempo para tomar algo?	1
Conozco un bar que está cerca, en esa esquina. Te invito una copa.	De acuerdo. Vamos.	1
Buenas tardes, ¿qué desean tomar?	Una cerveza para mí y un plato de patatas fritas para acompañar.	1
¿Algo más?	No, gracias. Es todo de momento.	1
¡Ah, me estoy muriendo de hambre! Comeré un bocadillo más. ¿Y tú, Roberto?	Estaba pensando que podemos aprovechar para merendar. Pediré yo también algo más. Pero yo te invito.	1
¡Camarera!	Un momento. Ahora voy ... ¿Sí?	1
Un momento. Ahora voy ... ¿Sí?	Quiero pedir un bocadillo de jamón y queso.	1
Quiero pedir un bocadillo de jamón y queso.	Para mí, otro igual pero con mucha lechuga.	1
¿Desean probarlo?	Sí, está bien.	1
Sí, está bien.	De acuerdo.	1
¿Tienes el periódico de hoy?	No. Lo siento. ¿por qué no lo compramos ahí?	1
No. Lo siento. ¿por qué no lo compramos ahí?	¡Qué bien!	1
¿Cuál es el periódico que se vende más en México?	Pues ... La Jornada.	1

Table A24: Examples of labeled data: Textbook (24)

¿Y por qué no los lees a través de Internet? Así podrías ver lo que más te guste de cada uno de ellos diariamente, ¿no?	Pues sí, a lo mejor tienes razón. Pero no es lo mismo mirar la pantalla de la computadora que pasar las páginas del periódico.	1
Pues sí, a lo mejor tienes razón. Pero no es lo mismo mirar la pantalla de la computadora que pasar las páginas del periódico.	Es verdad. Pues vamos a ver lo que hay en el periódico de hoy.	1
Yo voy a echar una mirada mientras a mi horóscopo.	No me digas que tú crees en los horóscopos ...	1
No me digas que tú crees en los horóscopos ...	No, no me interesa tanto lo que dicen, porque no lo creo, sino las cosas que inventan.	1
¿Qué raro! Normalmente es muy puntual.	No te preocupes. Seguro que no le ha pasado nada.	1
Voy a llamarle por teléfono.	Ya lo he intentado hace unos minutos. Por eso, si le llamas, no te contestaría.	1
Siento llegar retrasada ...	¿Qué te pasó? ¿Cómo es que llegas tan tarde?	1
Ya estábamos preocupados por ti.	Perdóneme.	1
Deberías haberme llamado ... Llevamos una hora esperándote.	Antes de salir de casa te llamé, pero tu teléfono estaba comunicando. Lo siento de verdad.	1
Estaría llamándote a ti. Pero ¿por qué no llamaste después? Tienes celular, ¿no?	Con las prisas, lo dejé en casa. Ya sabes, siempre estoy despistada. Lo siento de veras. Esto no volverá a suceder.	1
La semana pasada se estrenó una película de acción que me han dicho que está muy buena. ¿Te gustaría venir?	¿De verdad? Sí, sí tengo muchas ganas de ver una buena película. Hace tiempo que no voy.	1
¿En qué cine la ponen?	Pues, en varios.	1
Tú, ¿cuál prefieres ver, la versión original subtitulada o la doblada al español?	La original.	1
¿Qué te parece a las nueve?	Sí, me parece muy bien.	1
A propósito, tendremos que comprar las entradas en la taquilla, ¿no?	No, no hace falta. Voy a hacer la reservación por Internet.	1
Así no tendríamos que formarnos en la cola.	Muy bien.	1
Y, ¿quién actúa?	Antonio Banderas.	1
Si me hubieras dicho antes, habríamos invitado a Florencia, ya sabes que es fanática de Banderas.	Es cierto, lo olvidé.	1
Tal vez si le llamara ahora podría venir con nosotros.	Buena idea.	1
¿No me invitas un café?	Sí, claro. Pasa. De hecho, me iba a tomar uno ahora mismo.	1
Por cierto, ¿viste anoche el reportaje sobre el medio ambiente?	Sí, lo vi. Pero fue algo superficial, a tal punto que no explicaba nada sobre los problemas ambientales ni tampoco proponía soluciones. Tan solo quería impresionar a los espectadores.	1
Entonces, ¿qué tipo de programas te interesan?	Prefiero las telenovelas.	1
¿No hay manera de ver una película sin que la corten veinte veces?	Aquí, no.	1
¿Quieres que cambie de canal?	No tiene caso, de todos modos van a cortar cualquier programa bueno que encontremos.	1
Toma, un pedazo del pastel que te gusta. Para que estés contento ...	Tú sí conoces mi punto flaco. ¡Mmmm!	1
Te invito para que lo veamos juntos.	Gracias.	1
¿A qué hora nos vemos?	Comienza a las once quince, ¿te parece bien a las diez y media?	1
Así alcanzaremos a preparar algunos bocadillos.	Muy bien, a las diez treinta. ¡Los bocadillos son mi punto fuerte! Yo compraré lo necesario.	1
He comprado pan, queso, jamón, salami y, por supuesto, aceitunas con pimienta roja.	O sea, que has dejado vacía la tienda ...	1
Apuesto a que gana el equipo griego.	¿Sí? Pues ... y o apuesto al coreano, porque juega tan bien que hasta logró pasar a los cuartos de final en la Copa Mundial del 2002.	1
¿Qué apostamos?	Una cena en un buen restaurante que elija el ganador.	1
Una cena en un buen restaurante que elija el ganador.	De acuerdo. ¡Apostado!	1
¿Cómo va el partido? ¿Quién va ganando?	Corea está ganando uno a cero.	1
El árbitro acaba de marcar un penalti a favor de Corea.	¡Es Pak, lanza y ... gooooo! ¡Era imparable!	1
¿Adónde quieres ir a cenar?	Quisiera comer un buen filete, me gustaría cenar en "El rincón gaucho".	1
Entonces, ¿vemos otra vez juntos el partido de la final, pasado mañana?	¡Desde luego que no! Muchas gracias.	1
¡Señor agente! Vengo a denunciar un robo.	Tenga esta forma. Llénela con sus datos personales;	1
Aquí está.	Bien.	1
¿Algún objeto desaparecido?	Las maletas y el radio estéreo.	1
Las maletas y el radio estéreo.	¿Y qué tenían las maletas?	1

Table A25: Examples of labeled data: Textbook (25)

¿Y qué tenían las maletas?	Los pasaportes, 1.500 euros en cheques de viajero, una tarjeta de crédito y ropa.	1
¿Tiene seguro de viaje?	Sí. El de enfermedad y el de robo.	1
¿Podría recuperar los cheques de viajero?	Si tiene la factura, no hay problema.	1
Conviene que se dé prisa para que no utilicen su tarjeta.	Muchas gracias.	1
Muchas gracias.	De nada. Lo lamento, de verdad.	1
Buenos días, señorita. Tengo que mandar esta carta a París, ¿qué debo hacer?	Pase a aquella ventanilla, la que está al lado de la puerta K.	1
Debe llegar lo antes posible, así que tengo que mandarla como envío urgente.	Entonces, le recomiendo que haga el envío como carta certificada.	1
Para mandar una carta certificada hay que llenar el impreso que está en el mostrador.	¿Cómo lo debo llenar?	1
¿Cómo lo debo llenar?	Aquí escriba el nombre, apellido y dirección del destinatario, y donde dice remitente, ponga su nombre y sus datos.	1
Tiene que escribir con letra de imprenta.	De acuerdo, gracias.	1
Enseguida vuelvo.	Bien, todo perfecto.	1
Ahora ponga esta calia en el buzón amarillo, donde está el policía.	¡Ah, sí! Ya lo vi.	1
Muchas gracias, es usted muy amable.	Para servirle.	1
¿A cuánto está el dólar?	¿A la compra o a la venta?	1
¿A la compra o a la venta?	A la venta.	1
Me gustaría cambiar estos dólares por euros.	El dólar está a 0.82 euros.	1
¿Tiene cuenta en dólares en nuestro banco?	Me temo que no.	1
Me temo que no.	Entonces, le recomiendo que abra una.	1
Si tiene una cuenta corriente en nuestro banco, no le cobramos comisión por transferencias y, además, le ofrecemos un tipo de cambio más favorable.	¿Sí? Entonces, haré como me sugiere y abriré una cuenta en dólares.	1
¿Quiere una cuenta corriente o de ahorros?	Prefiero una cuenta corriente.	1
Prefiero una cuenta corriente.	Bien. Llene este formulario y firme aquí.	1
¿Cuánto va a depositar?	Cuatro mil trescientos dólares.	1
¿Puedo hacer una transferencia al extranjero a través de esta cuenta?	Claro que sí, y no le cobramos comisión, como le dije antes.	1
Si su esposo desea firmar también, puede venir a nuestras oficinas y registraremos su firma.	¡Perfecto! Se lo diré.	1
Le agradezco su amabilidad.	Estamos para servirle.	1
¿Me puede echar una mano?	Claro que sí.	1
Si quiere, le dejo algo de gasolina que tengo en el maletero de mi auto.	Sí, le agradezco.	1
¿Funciona ahora?	No, no arranca todavía.	1
No, no arranca todavía.	Si no funciona, le aconsejo llevarlo al mecánico ya que pronto anochecerá.	1
¿Podría ayudarme a empujar mi coche hasta allá? Disculpe, estoy abusando de usted, pero si quisiera llevarlo al taller mecánico.	No se preocupe.	1
¡Ha sido tan amable! Muchísimas gracias.	De nada. Ha sido un placer poder ayudarla. Adiós.	1
Mire, señor, se paró mi coche en medio de la carretera.	Espere un momento, lo veré en seguida.	1
Hace un ruido extraño.	¿Es el motor? ¿Estará roto?	1
¿Es el motor? ¿Estará roto?	Creo que no. Pero es mejor que se lo revisen de nuevo.	1
Hay poco tráfico aunque es fin de semana.	Normalmente, el tráfico está menos congestionado en septiembre que en julio y agosto.	1
¿No crees que sea mejor tomar la autopista que la carretera local para ir a los Pirineos?	Sí, creo que sí. Pero a veces no merece la pena utilizar la autopista.	1
Aunque podríamos ahorrar unos minutos, la cuota es muy cara.	Sí, es verdad.	1
Si tuviéramos más tiempo, estaría bien pasar por los caminos rurales y contemplar el paisaje.	Pues, no sé.	1
Y aun cuando la otra vía es más rápida, iremos por los caminos rurales para disfrutar el paisaje.	Lo disfrutaremos, ya verás.	1
A tres kilómetros de aquí se encuentra un restaurante típico que me recomendaron.	Perfecto. Podríamos comer allí.	1

Table A26: Examples of labeled data: Textbook (26)

Perfecto. Podríamos comer allí.	A pesar de estar lejos de la ciudad, tiene todo lo necesario para ofrecer un banquete a sus clientes.	1
¿Qué vas a hacer el sábado? ¿Vas a salir?	Sí, me gustaría dar una vuelta por el centro antiguo.	1
Sí, me gustaría dar una vuelta por el centro antiguo.	Si quieres ir al centro antiguo, te aconsejo que vayas a pie o en metro.	1
Si quieres ir al centro antiguo, te aconsejo que vayas a pie o en metro.	¿Y eso por qué?	1
¿Y eso por qué?	Porque anunciaron para mañana huelga de transporte y no habrá autobuses.	1
Se irán a huelga a fin de conseguir mejores salarios.	Entonces, tendré que tomar el metro.	1
¿Cuál es la estación del metro más cercana de aquí?	La estación Cuatro Caminos.	1
¿De acuerdo?	Gracias. Pero ¿dónde puedo comprar el billete del metro?	1
Gracias. Pero ¿dónde puedo comprar el billete del metro?	En la taquilla o en la máquina expendedora.	1
El bono te sale mucho más barato.	Gracias por la información. Compraré el bono para ahorrar unos euros.	1
Pase y tome asiento.	Gracias.	1
Creo que llego con un poco de retraso. Lo siento, el tráfico estaba terrible.	No se preocupe, es algo que le puede pasar a cualquiera.	1
Dígame, ¿dónde realizó sus estudios universitarios?	Estudié la carrera de Economía en la Universidad Politécnica y me especialicé en Ciencias Económicas y Empresariales.	1
Al terminar la carrera, cursé la Maestría en Dirección y Administración de Empresas, en los Estados Unidos.	¿Tiene alguna experiencia en este campo?	1
¿Tiene alguna experiencia en este campo?	Sí.	1
Trabajé en una empresa de importación y exportación durante dos años después de graduarme.	Como sabe usted, para este puesto estamos solicitando algunos requisitos.	1
Como sabe usted, para este puesto estamos solicitando algunos requisitos.	Creo que yo sí cumpla con todos los requisitos. Domino bien el inglés y el francés.	1
¿Tiene usted disponibilidad para viajar?	Sí, absolutamente.	1
Muy bien, por hoy ya ha terminado conmigo. Nos veremos mañana.	Gracias por su tiempo, hasta mañana.	1
Pase con ellos y espere su turno con el Director de personal.	Está bien. Gracias de nuevo.	1
¿Podría recibirme el dentista?	¿Tiene cita para hoy?	1
¿Tiene cita para hoy?	No. Pero ¿podría recibirme sin cita?	1
Necesito que el doctor me revise de urgencia.	¿Le importaría esperar un momento? Veré qué puedo hacer por usted.	1
¿Le importaría esperar un momento? Veré qué puedo hacer por usted.	Esperaré. No tengo otra opción. Y le agradezco.	1
¿Quiere pasar a la sala de consulta y sentarse, por favor?	Gracias, señoría.	1
Dígame, ¿Qué problema tiene?	Tengo molestias en la muela.	1
Tengo molestias en la muela.	Vamos a ver... Abra la boca, por favor.	1
Como usted crea conveniente, doctor.	Le voy a poner un poco de anestesia.	1
Ahora dígame, ¿cómo se siente?	Mucho mejor, doctor. Gracias.	1
¿Tengo que volver la próxima semana?	Sí. Pida una cita a la enfermera.	1
¿Es usted alérgico a alguna medicina?	No, no soy alérgico a ninguna medicina...	1
Perdón, ahora recuerdo que sí soy alérgico a la penicilina.	Doctora: Bien. ¿Y qué le pasó?	1
Cuénteme, ¿cómo se ha hecho la herida?	Me caí de una escalera.	1
¿Cuánto tiempo hace que le pusieron una inyección antitetánica?	No recuerdo exactamente.	1
Entonces, pregunte a quien le puso la vacuna. Luego me avisa.	Está bien.	1
Doctora, me duele también el tobillo. Creo que lo tengo roto o torcido.	Vamos a hacerle una radiografía. Temo que haya una fractura.	1
Hummm... desgraciadamente, usted tiene el tobillo fracturado. Voy a ponerle una férula.	Muy bien, doctora. Ojalá que no sea muy doloroso. No me gusta tomar calmantes.	1
Muy bien, doctora. Ojalá que no sea muy doloroso. No me gusta tomar calmantes.	Le daré un analgésico suave que no dañará su organismo. Aquí tiene la orden.	1
¿Tiene usted seguro médico?	Sí, y espero que el seguro cubra todos los gastos.	1
Sí, y espero que el seguro cubra todos los gastos.	Entregue esta constancia a quien le vendió el seguro. Con esta constancia le pagarán todos los gastos.	1
Nos veremos en tres días.	Y aquí nos veremos en tres días.	1
¿Cuándo comenzó a sentirse mal?	Desde anteayer.	1
Me pasé dos días vomitando. Además, tengo diarrea y náuseas.	¿Qué otros síntomas tiene?	1
¿Qué otros síntomas tiene?	Me siento muy débil. No tengo fuerzas y me duele la espalda.	1
¿Hay alguna persona diabética en su familia?	No. Yo soy la única diabética en mi familia.	1
Y me preocupa mucho porque estoy embarazada.	No se preocupe...	1
Ahora respire hondo. Otra vez.	Me duele el pecho cuando respiro y también me duelen los oídos.	1
¿Y para el dolor de oído, doctor?	Voy a recetarle un antibiótico que no afecta al feto.	1
No es bueno que coma mariscos ni pescado en este periodo.	El pescado que comí anteayer estaba muy fresco.	1
Vuelva mañana. Necesito hacerle unos análisis que complementen el diagnóstico.	Muy bien, doctor. Muchas gracias.	1

Table A27: Examples of labeled data: Textbook (27)

¡Hola!	¿Cómo te llamas?	2
Mi nombre es Pedro García.	Encantada de conocerte.	2
Se escribe L-O-P-E-Z.	¡Muchas gracias!	2
¡De nada!	¡Hasta otro día!	2
Kim.	¿Cómo se escribe?	2
Mi nombre es Raúl.	¿Cómo se escribe su nombre?	2
¡Genial!	¿Qué tal tu clase de español?	2
Azules, blancas, rojas y amarillas.	¡Sumi! ¿Qué grande es tu mochila? ¿Qué hay?	2
¡Buenos días!	Mira, te presento a mi mejor amigo.	2
Encantado.	¿Eres español?	2
Soy Luis Carlos González Pérez.	¿Cuál es su profesión?	2
Soy ingeniero electrónico.	¿Cuántos idiomas habla?	2
Hablo tres idiomas: español, francés e inglés.	¿Cuáles son sus cualidades?	2
Soy trabajador, sincero, entusiasta, amable y puntual.	¿Cuál es su nacionalidad?	2
Soy chileno.	Muchas gracias por su tiempo, Sr. Pérez.	2
Soy Pedro, el profesor de español.	¿Quiénes son ellos?	2
Muchas gracias, señora Pérez.	¿Qué tal el viaje?	2
¡Claro que sí! Hay una farmacia muy grande enfrente de la casa.	Y la universidad, ¿dónde está?	2
La universidad está un poco lejos, pero hay una parada del bus universitario delante de la casa.	Una última pregunta. ¿Hay un banco por aquí?	2
Hay muchos bancos dentro del centro comercial del barrio. El centro comercial está entre la oficina de correos y la plaza.	¿Qué bien! Mil gracias, señora Pérez.	2
Me llamo Mison.	¿Dónde vives, Mison?	2
Vivo en el centro de la ciudad en una pensión familiar.	¿Qué bien hablas español!	2
¡Hola, Carlos!	¿Estás ocupada este fin de semana?	2
Sí, mañana es el cumpleaños de mi abuelo.	¿Cuántos años cumple tu abuelo?	2
Setenta y cinco años. Mi abuela también tiene la misma edad.	¿Dónde viven tus abuelos?	2
Ellos viven en Córdoba.	¿Viven solos tus abuelos?	2
A ver ...	Mira, estos del centro son mis abuelos y a su derecha están mis padres. Esta es mi hermana Carolina y este soy yo.	2
Gracias.	¿Qué hacen tus padres?	2
Mi padre es diplomático y mi madre es ama de casa.	¿Vives con ellos?	2
¿Sí?	Pensamos hacer una "fiesta sorpresa" en su casa. ¿Puedes asistir?	2
Pero mañana jueves voy al centro comercial para comprar un regalo.	Bueno, entonces nos vemos pasado mañana en el club. De allí vamos todos juntos a casa de Carla.	2
¡Qué emoción! Sois muy buenos amigos.	¡Ahora la tarta y las velas! Uno, dos, tres... Cumpleaños feliz, cumpleaños feliz, te deseamos todos, cumpleaños feliz ¡Bravo! (aplausos)	2
Empieza a las diez de la mañana y termina a las once y media.	¿A qué hora abren los bancos?	2
Abren a las nueve.	¿A qué hora cierran?	2
Es el 2 de marzo.	Entonces, ¿a cuántos estamos?	2
Estamos a primero de marzo.	¡Feliz cumpleaños!	2
No sé todavía. Como es un viaje para mis padres, quiero elegir un buen lugar, sobre todo, un lugar con buen clima.	¿Qué le parece Sudamérica?	2
Hace mucho calor y llueve mucho. El clima es muy húmedo.	¿Y en invierno?	2
En el cine que está enfrente del ayuntamiento.	¿A qué hora y dónde quedamos?	2
Sí, son emocionantes.	Aparte del fútbol, ¿qué otras cosas te gustan?	2

Table A28: Examples of labeled data: Textbook (28)

Un momento por favor. Ahora se pone.	Hola. Clara. ¿Qué hay?	2
Una falda corta.	¿De qué estilo?	2
Una falda vaquera de algodón.	¿De qué color?	2
Quiero una falda de color azul claro.	¿De qué talla?	2
No sé-- una falda de talla mediana, o un poco más pequeña.	Tiene que probársela antes de comprarla. Cada marca tiene tallas diferentes.	2
¡Es verdad! También, según el diseño, a veces me va bien una talla más pequeña.	¿La quiere con bolsillos?	2
Cuesta cien euros.	En total, ¿cuánto es?	2
¡Ya voy mamá!...	Con una taza de leche caliente y una cucharada de miel se te pasa.	2
¡Mamá! ¡Solo paso treinta minutos en el baño!	¡Date prisa! Tienes que desayunar y no te queda mucho tiempo.	2
¡No te preocupes, mamá! Tengo tiempo suficiente.	No olvides que, después de desayunar, debes cepillarte los dientes.	2
Yo creo que Madrid y Barcelona son las ciudades que debes visitar.	Y ¿has probado la comida española?	2
¡Buenas tardes!	¿A qué hora sale el próximo tren a Sevilla?	2
Aquí tiene.	¡Muy bien!	2
¡Es mi poeta favorito!	¿Y por la noche qué hicieron?	2
Fuimos a Granada y cenamos en un restaurante árabe.	¿Visitaron La Alhambra?	2
¡Vaya! Eso sí fue un buen fin de semana.	¿Y tú adónde fuiste?	2
Mi padre practicaba deportes acuáticos y durante las vacaciones solía llevarme al mar para nadar y pescar.	¿Qué bien recuerdas tu infancia!	2
Iré en avión a Cancún y pasaré allá un fin de semana fantástico. Me alojaré en un hotel muy bonito frente al mar y, por la noche, beberé y bailaré en alguna discoteca.	¿Bailas bien?	2
Soy ecuatoriana y nací bailando ...	Para mí, estas serán mis últimas vacaciones de soltera, así que debo aprovecharlas antes de volver a Ecuador.	2
Sí, para dos personas.	¿Les parece bien esta?	2
Muy bien, gracias.	¿Desearían comer algo?	2
Para mí, un helado de chocolate.	¿Quieren café o té?	2
Un café con leche, para mí.	¿Desean algo más?	2
Bien, bien, entiendo...	Siga caminando por el Paseo del Prado, unas cuatro calles, luego gire a la izquierda hacia la Calle Felipe IV. El Museo del Prado está exactamente a la derecha de esta calle.	2
¡Carlos, soy María! Te llamo desde una cabina de la Plaza de España.	¿Pasa algo? ¿Estás bien?	2
Mi móvil, mi cartera con dinero, tarjetas de débito y crédito, mi DNI, un moneo pequeño y otras cosas más.	Te recomiendo que canceles inmediatamente tus tarjetas. Luego, iremos a una comisaría para denunciar el robo. ¡Ahora mismo voy!	2
¡Espero que sea así!	Después, hay que ir rápidamente a una comisaría.	2
No te preocupes. Mañana conduciré yo. Solo espero que haga buen tiempo...	En el periódico he visto que mañana hará sol.	2
Descuida. Mejor dámela ahora y la meto en la maleta. ¿Y qué vamos a hacer en Toledo?	Mira, hay numerosas opciones para recorrer la ciudad. Hay rutas históricas, artísticas y también existen las rutas de don Quijote donde se pueden ver algunos parajes que aparecen en El Quijote.	2
Genial, siempre he querido visitar uno ... Me encanta el capítulo de los molinos de viento de El Quijote.	¡Ah! En cuanto lleguemos al hotel, pregunta qué autopista debemos tomar para ir a Toledo.	2
Ayudo a la gente mayor que vive en una residencia de la tercera edad.	Pues, estando solo con ancianos, no creo que el voluntariado sea divertido ...	2
Al contrario, me encanta ir. Disfruto mucho conversando con ellos.	¿Y de qué hablas con ellos que pueda ser interesante?	2
Pues a mí me parece que han hecho un buen trabajo.	No puedo creer que hayan prohibido las corridas de toros en Cataluña.	2
Pues yo me alegro de que el gobierno lo haya decidido así.	Se trata de una tradición muy antigua, no sé qué tiene de malo.	2
Es una pena que de un tiempo acá la gente haya visto las corridas como actos de violencia, porque más bien se trata de un arte, de un ritual.	No entiendo cómo pueden considerarlas manifestaciones artísticas cuando son una práctica de crueldad contra los toros.	2
Ah, sí ... Me alegré de verla porque cuando llegué no había nadie que me conociera.	Me dijo que buscaban a alguien que supiera hablar muchos idiomas.	2
Sí. Ese es el problema, solo hablo inglés y un poco de alemán.	No te desanimes. Aún no conoces los resultados.	2
Si confiaras más en ti mismo, las cosas te irían mejor.	Además, uno de los entrevistadores era un antiguo profesor que tuve en el primer año. ¿Y sabes qué? Me miró como si no me conociera.	2
Hombre, siendo entrevistador es lógico que tuviera que ser objetivo.	Sí, pero luego lo volví a ver en el pasillo y me ignoró, pasó como si yo no existiera.	2
¿Mi hermana? ¿Por qué?	Ya sabes que me gusta desde hace tiempo ... y si pudiera verla de nuevo ... te lo agradecería infinitamente.	2

Table A29: Examples of labeled data: Textbook (29)

Ah, si ... Me alegré de verla porque cuando llegué no había nadie que me conociera.	Me dijo que buscaban a alguien que supiera hablar muchos idiomas.	2
Si. Ese es el problema, solo hablo inglés y un poco de alemán.	No te desanimes. Aún no conoces los resultados.	2
Si confiaras más en ti mismo, las cosas te irían mejor.	Además, uno de los entrevistadores era un antiguo profesor que tuve en el primer año. ¿Y sabes qué? Me miró como si no me conociera.	2
Hombre, siendo entrevistador es lógico que tuviera que ser objetivo.	Si, pero luego lo volví a ver en el pasillo y me ignoró, pasó como si yo no existiera.	2
¿Mi hermana? ¿Por qué?	Ya sabes que me gusta desde hace tiempo... y si pudiera verla de nuevo ... te lo agradecería infinitamente.	2
Si no me hubieras llamado por mi nombre, no la habría reconocido.	¿Ha cambiado mucho?	2
Ha perdido mucho peso, yo diría que al menos unos diez kilos, está más guapa.	Si me hubieras llamado por teléfono, habría bajado al parque para verla.	2
Si me hubieras llamado por teléfono, habría bajado al parque para verla.	Ella tenía mucha prisa. Si hubiera tenido más tiempo, le habría pedido que subiera a casa para que os saludarais.	2
Ella tenía mucha prisa. Si hubiera tenido más tiempo, le habría pedido que subiera a casa para que os saludarais.	¿Qué lástima! Hace tantos años ... , hubiera querido darle un abrazo.	2
¿Y qué más te contó?	Tranquila, me pidió tu número. Te llamará un día de estos.	2
Pues, que se dedicaba a adiestrar perros de asistencia.	Pues, que se dedicaba a adiestrar perros de asistencia.	2
Son perros especialmente adiestrados para ayudar a personas con algún tipo de discapacidad o que hubieran sufrido un accidente.	¿Qué son los perros de asistencia?	2
Muy bien. He estado en mi pueblo durante todas las vacaciones y disfruté como si hubiera vivido siempre allí.	No sabía que Alicia hubiera hecho un curso de adiestramiento para perros.	2
Vaya, ¡qué envidia! Yo me quedé todo el verano por aquí ... un aburrimiento.	Vaya, ¡qué envidia! Yo me quedé todo el verano por aquí ... un aburrimiento.	2
¿Y qué has venido a hacer hoy a la facultad? No creo que las clases empiecen hoy sino hasta la semana que viene.	¿Y qué has venido a hacer hoy a la facultad? No creo que las clases empiecen hoy sino hasta la semana que viene.	2
Cierto, solo que estoy mirando si quedaran plazas en la residencia de la universidad.	Cierto, solo que estoy mirando si quedaran plazas en la residencia de la universidad.	2
Si, pero ojalá que pudiera dejar el piso.	¿No estabas viviendo en un piso compartido cerca de aquí?	2
Si, pero con tanta gente es imposible estudiar en casa y muchas veces no tengo tiempo para cocinar ni para hacer las tareas del hogar.	¿Por qué? Decías que era muy económico y bien ubicado.	2
Lo sé, ojalá tenga suerte y pueda entrar a vivir en la residencia.	Entiendo. Pero a estas alturas puede que sea difícil encontrar una habitación disponible.	2
¿De qué se trata?	Hablas como si se tratara de la lotería ...	2
¿Cómo podrían ser un problema?	Habla de la creciente inquietud en la red por los continuos problemas en el servicio de mensajería móvil gratuita.	2
Porque la aplicación universal de mensajería por Internet ha aumentado en todo el mundo.	Porque la aplicación universal de mensajería por Internet ha aumentado en todo el mundo.	2
¿Y quién sale perjudicado?	¿Y quién sale perjudicado?	2
Pues, los operadores de telefonía.	Pues, los operadores de telefonía.	2
Dice que los ingresos por concepto de mensajes han descendido mucho.	A ver ... , ¿de dónde es la noticia?	2
¡Claro! Mira, si las tarifas que ofrecieran fueran más baratas, la gente no dependería tanto de los mensajes gratuitos. A veces no entiendo cómo las condiciones de los contratos son tan abusivas.	¡Claro! Mira, si las tarifas que ofrecieran fueran más baratas, la gente no dependería tanto de los mensajes gratuitos. A veces no entiendo cómo las condiciones de los contratos son tan abusivas.	2
Es cierto, pero gracias a la tecnología estamos comunicados durante las 24 horas del día y eso es muy importante en casos de emergencia.	Es cierto, pero gracias a la tecnología estamos comunicados durante las 24 horas del día y eso es muy importante en casos de emergencia.	2
prefiero quedarme aquí.	Estoy de acuerdo, pero a veces dependemos demasiado del móvil, ¿no crees?	2
Pues estoy escribiendo una carta para la familia que conocí en Bolivia.	¿Y qué escribes?	2
¿Y por qué no les envías un correo electrónico?	¿Y por qué no les envías un correo electrónico?	2
Se trata de una familia que conocí en Totorá, una pequeña ciudad que está al este de Cochabamba. La mayoría de las viviendas no tiene computadora ni tampoco dispone de conexión a Internet.	Se trata de una familia que conocí en Totorá, una pequeña ciudad que está al este de Cochabamba. La mayoría de las viviendas no tiene computadora ni tampoco dispone de conexión a Internet.	2
Bueno, aún no me has contado cómo fue el viaje por América Latina.	Bueno, aún no me has contado cómo fue el viaje por América Latina.	2
El otro día estaba leyendo que no era un país muy visitado por extranjeros.	El otro día estaba leyendo que no era un país muy visitado por extranjeros.	2
¿Y cuál fue tu paisaje favorito? Conociéndote, ¡debes haber tomado cientos de fotos!	¿Y cuál fue tu paisaje favorito? Conociéndote, ¡debes haber tomado cientos de fotos!	2
¿Qué hiciste el fin de semana?	¿Qué hiciste el fin de semana?	2
¿Qué fin de semana tan movido, ¿no?	¿Qué fin de semana tan movido, ¿no?	2
Nunca he acudido a ese tipo de exposiciones ... Prefiero el cine, el teatro, las modas ...	Nunca he acudido a ese tipo de exposiciones ... Prefiero el cine, el teatro, las modas ...	2
Un día iremos juntas o ... ¡mejor invita a tu novio!	Un día iremos juntas o ... ¡mejor invita a tu novio!	2
Ya verás cómo te gustará.	Ya verás cómo te gustará.	2
¿Vamos a tomar un café? Creo que la cafetería de ahí aún está abierta.	¿Vamos a tomar un café? Creo que la cafetería de ahí aún está abierta.	2

Table A30: Examples of labeled data: Textbook (30)

Ultimamente las cosas con Miguel no van muy bien.	¿Habéis discutido?	2
Apenas hablamos.	Es importante hablar cara a cara para solucionar los problemas.	2
Hace cosas que antes no solía hacer.	¿Qué quieres decir? ¿Que te engaña?	2
Pero a veces me da la sensación de que me evita.	Lo mejor es confiar en tu pareja, así que no saques conclusiones antes de tiempo.	2
Lo mejor es confiar en tu pareja, así que no saques conclusiones antes de tiempo.	Por ejemplo, ayer prometió venir a la cena de mi empresa y no apareció. Y de repente, todos los días hace deporte para no engordar.	2
Por ejemplo, ayer prometió venir a la cena de mi empresa y no apareció. Y de repente, todos los días hace deporte para no engordar.	Está bien que se quiera mantener en forma. Eso no significa nada. Yo solo te puedo aconsejar que hables con él para preguntarle directamente qué le pasa. La confianza es lo más importante en una relación.	2
Está bien que se quiera mantener en forma. Eso no significa nada. Yo solo te puedo aconsejar que hables con él para preguntarle directamente qué le pasa. La confianza es lo más importante en una relación.	Perdón por molestarte con mis problemas.	2
¿Cómo van los preparativos de la fiesta de quinceañera de tu prima María?	¡Genial! Está casi todo listo, solo quedan algunos retoques que se harán el mismo día.	2
¡Genial! Está casi todo listo, solo quedan algunos retoques que se harán el mismo día.	También van a venir vuestros tíos, los que viven en los Estados Unidos, ¿verdad?	2
También van a venir vuestros tíos, los que viven en los Estados Unidos, ¿verdad?	Si, ¡qué ilusión! Hace muchísimo que no los veo.	2
Si, ¡qué ilusión! Hace muchísimo que no los veo.	Y cuéntame.	2
Es ideal. Dispone de un jardín hermoso con un lago romántico.	Parece que va a ser una celebración por todo lo alto ... No sabía que este evento tuviera tanta importancia en vuestra cultura.	2
Parece que va a ser una celebración por todo lo alto ... No sabía que este evento tuviera tanta importancia en vuestra cultura.	Pues sí. Se trata de un rito de paso. La quinceañera que era considerada una niña hasta los catorce años, se presenta en sociedad como una mujer.	2
Pues sí. Se trata de un rito de paso. La quinceañera que era considerada una niña hasta los catorce años, se presenta en sociedad como una mujer.	¿Y qué parte de la celebración te gusta más?	2
Cambia sus zapatos planos por unos de tacón.	¿Qué interesante! ¿Y cuánto dinero se suele gastar en una celebración de este tipo?	2
¿Qué interesante! ¿Y cuánto dinero se suele gastar en una celebración de este tipo?	Eso depende de cada fiesta, pero el gasto promedio es entre unos 15 y 18 mil dólares.	2
Eso depende de cada fiesta, pero el gasto promedio es entre unos 15 y 18 mil dólares.	¡Madre mía! Mucho más de lo que imaginaba.	2
¿Cuándo volviste del pueblo? Pensé que no llegarías hasta la semana que viene.	Esa era mi intención, pero ha surgido un problema en el proyecto en el que estoy trabajando y he tenido que volver antes de lo previsto.	2
Esa era mi intención, pero ha surgido un problema en el proyecto en el que estoy trabajando y he tenido que volver antes de lo previsto.	Vaya, ¡qué mala suerte! Primera vez que te daban dos semanas de vacaciones ...	2
Vaya, ¡qué mala suerte! Primera vez que te daban dos semanas de vacaciones ...	¿Qué le vamos a hacer! Ya sabes, el trabajo es lo primero.	2
¿Qué le vamos a hacer! Ya sabes, el trabajo es lo primero.	Bueno, ¿cómo está la gente del pueblo? Hace tiempo que no sé nada de ellos.	2
Bueno, ¿cómo está la gente del pueblo? Hace tiempo que no sé nada de ellos.	¿Te acuerdas de Ana, la chica cuyo padre tenía una zapatería cerca de la plaza?	2
Pues se ha casado.	Me alegro por ella ... me extrañaba que aún no estuviera casada. Es una chica espléndida, en la que se puede confiar.	2
Me alegro por ella ... me extrañaba que aún no estuviera casada. Es una chica espléndida, en la que se puede confiar.	¿Y lo vuestro por qué no funcionó?	2
¿Y lo vuestro por qué no funcionó?	En aquel entonces éramos unos críos ... tan solo teníamos 17 años.	2
En aquel entonces éramos unos críos ... tan solo teníamos 17 años.	Oye, la próxima vez tenemos que ir juntos.	2
¿Que querías preguntarme antes?	Era acerca del trabajo sobre los pueblos indígenas de Honduras.	2
De hecho, en el ordenador de casa tengo varios ensayos sobre el tema.	Pero, ¿tenías el ordenador estropeado, no es verdad?	2
Si, pero el problema fue detectado por un técnico y ya está solucionado.	Menos mal. Oye, ¿sabes cuántos pueblos indígenas hay en Honduras?	2
Menos mal. Oye, ¿sabes cuántos pueblos indígenas hay en Honduras?	Se reconoce la presencia de nueve pueblos culturalmente diferenciados.	2
Se reconoce la presencia de nueve pueblos culturalmente diferenciados.	¿Sabes dónde se encuentran localizados?	2
El área geográfica de influencia de estos pueblos indígenas está distribuida en todo el país, específicamente por el occidente, el sur y la zona central.	¿Tienes datos sobre la población?	2
Ya que la información correspondía a censos de hace más de 20 años. Pero los porcentajes con relación a la población nacional van del 7 al 13 por ciento.	Hay una exposición sobre pueblos indígenas de América Latina.	2
Pero tanto si vienen como si no, nosotros vamos, ¿eh?	Pues sí que te gusta ir a ese restaurante, no lo hubiera imaginado ...	2
Pues sí que te gusta ir a ese restaurante, no lo hubiera imaginado ...	No es eso, más que nada me apetece salir de casa ...	2
No es eso, más que nada me apetece salir de casa ...	¿Sabías que en una de las salas venden piezas de artesanía de Chile?	2
¿Qué es eso?	Una piedra semipreciosa de color azul procedente de la precordillera de la región de Coquimbo. El trabajo con esta piedra se luce en joyas y adornos con figuras de animales, jarrones y mosaicos.	2
Una piedra semipreciosa de color azul procedente de la precordillera de la región de Coquimbo. El trabajo con esta piedra se luce en joyas y adornos con figuras de animales, jarrones y mosaicos.	¿Qué interesante!	2
Pues no lo es. Que yo sepa, no tiene más de diez años.	Oye, antes de seguir hablando de otras cosas, más vale que llamemos para reservar una mesa.	2
La tarifa es de setenta euros y ustedes pueden estar todo el tiempo que quieran sin otro límite que el horario de la mezquita.	¿Y cómo son las visitas en grupo?	2

Table A31: Examples of labeled data: Textbook (31)

Además, pueden elegir el horario que más les convenga.	Un momento, vamos a discutirlo y ahora volvemos.	2
Un momento, vamos a discutirlo y ahora volvemos.	Dado que hoy ya no nos queda mucho dinero, no tenemos otro remedio que hacer la visita en grupo.	2
Dado que hoy ya no nos queda mucho dinero, no tenemos otro remedio que hacer la visita en grupo.	Pero, ¿podemos pagar con la tarjeta de crédito, no?	2
¿Y si volvemos mañana?	Mira, lo del precio es lo de menos.	2
Roberto, nada más te quejas ...	Mi idea de hacer un viaje no consiste tan solo en visitar monumentos y museos. No hemos hecho otra cosa que ir a palacios y exhibiciones ...	2
Mi idea de hacer un viaje no consiste tan solo en visitar monumentos y museos. No hemos hecho otra cosa que ir a palacios y exhibiciones ...	Vaya, no sabía que te estabas aburriendo tanto ...	2
Ayer se quedó estudiando hasta la madrugada y esta mañana se ha levantado tarde.	Por cierto.	2
Creo que se lo leyó completo el fin de semana.	Últimamente está muy estudioso. ¿Qué mosca le habrá picado?	2
Últimamente está muy estudioso. ¿Qué mosca le habrá picado?	Pues que por fin abrió los ojos. Se ha dado cuenta de lo mal que está el mercado laboral. Así, que ha decidido ponerse las pilas ... aunque lo debería haber hecho antes.	2
Pues que por fin abrió los ojos. Se ha dado cuenta de lo mal que está el mercado laboral. Así, que ha decidido ponerse las pilas ... aunque lo debería haber hecho antes.	Bueno, mejor tarde que nunca, ¿no?	2
Se me ocurren mil cosas. No sé qué hacer con mi vida. Por eso, estoy pensando en recorrer el Camino de Santiago.	¿Y eso? Pensaba que no te gustaban los viajes sacrificados.	2
¿Y eso? Pensaba que no te gustaban los viajes sacrificados.	Necesito un respiro ... Tengo que ordenar mis ideas.	2
Necesito un respiro ... Tengo que ordenar mis ideas.	Pero, ¿no estabas pensando en buscar un trabajo?	2
Con la crisis, la situación se ha puesto muy mal.	¿Sabes que se ha convocado a un concurso de plazas en el Ministerio de Cultura?	2
¿Sabes que se ha convocado a un concurso de plazas en el Ministerio de Cultura?	Sí, pero solo hay una. Además, se requiere el conocimiento de por lo menos, dos lenguas extranjeras. Por eso, también estaba pensando en ir al extranjero para aprender idiomas.	2
Sí, pero solo hay una. Además, se requiere el conocimiento de por lo menos, dos lenguas extranjeras. Por eso, también estaba pensando en ir al extranjero para aprender idiomas.	¿Y cómo te pagarías los gastos si no trabajas?	2
Me gustaría ir a un país donde se hable francés, pero también donde se pueda practicar el inglés.	Eso lo puedes decidir mientras haces el camino ...	2
¿Qué te pasa? Te noto un poco nerviosa.	¡Qué voy a hacer!	2
Bueno, no lo soy. Pero mi madre siempre decía: "¡Más vale no tentar a la suerte!".	Y hoy la necesitas más que nunca, ¿no?	2
Tengo que aprobar el último examen de la carrera. Por si las moscas, llevo mi trébol de cuatro hojas.	Luego dices que no eres supersticiosa ... Bueno, tú, tranquila.	2
Ayer estuve hablando con mis padres y me dijeron que si no aprobaba este examen dejarían de darme mi paga mensual y que debería buscar un trabajo a tiempo parcial para pagar mis gastos.	Sí que te presionan ... Y yo que te dije que no tendrías ningún problema ...	2
Sí que te presionan ... Y yo que te dije que no tendrías ningún problema ...	Están deseosos de que me licencie y que busque un trabajo lo antes posible.	2
Ayer, precisamente, quedé con él para tomar algo. Me dijo que estaba pensando en buscar otro trabajo.	¿Por qué? Lleva menos de un año en esa empresa.	2
Además, que con el sueldo que recibía no llegaba al final del mes.	¡Vaya! Y yo sin saber nada de eso. Justamente ayer le dejé un mensaje en el móvil: "Victor, ¡a ver si me invitas a cenar ahora que trabajas en una multinacional!" ¡Qué metedura de pata!	2
¿Qué va! Si me ha encantado, ya sabes que me interesa muchísimo este tipo de películas.	¿Entonces? ¿Por qué estás enfadada?	2
Ah, ahora entiendo. Lo dices por el chico que estaba sentado a tu lado. Le ha sonado el teléfono unas cuantas veces.	Un par de veces no me importa, pero estuve enviando mensajes durante toda la película.	2
Me parece bien, me agrada mucho ese restaurante.	Pero tendremos que comer rápido.	2
Bueno, aun nos queda una hora y media.	Ya sabes que ese restaurante siempre está lleno, espero que no nos falte tiempo para llegar a la hora.	2
Pues dondequiera, en la papelería si gustas.	¿Con los papeles?	2
Cuando se llene, lo tiraré todo junto.	Maria, no me digas que a estas alturas aún no clasificas la basura cuando la tiras.	2
De hecho, no conozco a nadie que lo haga.	Menuda excusa tienes para no reciclar. Millones de personas que piensan como tú agravan la contaminación atmosférica.	2
Menuda excusa tienes para no reciclar. Millones de personas que piensan como tú agravan la contaminación atmosférica.	Ahora resulta que yo soy la culpable de los problemas del medio ambiente!	2
No te lo tomes a broma. Es un tema muy grave y alguien tiene que ponerle remedio.	Por el hecho de que yo recicle, el medio ambiente no va a mejorar.	2
Por el hecho de que yo recicle, el medio ambiente no va a mejorar.	Cualquier acción, por pequeña que sea, contribuirá a mejorar este planeta.	2
Bueno, no sabía que estuvieras a favor del movimiento ecológico.	No defendiendo ningún movimiento, solo pienso que es responsabilidad de todos actuar correctamente para preservar nuestro entorno.	2
Buenas tardes, ¿en qué puedo servirle?	Buenas tardes, ¿es esta la residencia de la escuela de idiomas?	2
Sí, aquí es.	Me inscribí por Internet. Mi nombre es Carolina Hernández.	2
Che (de chico), o (de Oscar), i (de iglesia).	Un momento por favor..., aquí está. Estaba en la lista de pendientes.	2
Cero. Cero, cuarenta y ocho, setenta y cinco, diez.	Bien.	2
La que usted pagó era compartida. Su compañera llegó esta mañana.	Muchas gracias, muy amable.	2

Table A32: Examples of labeled data: Textbook (32)

Lee.	¿Apellido materno?	2
No aparece en mi pasaporte. Pero es kim.	No importa, lo necesitamos en este formulario.	2
El 16 de mayo de 1983, en Seúl.	¿Nacionalidad? ¿Estado civil?	2
Coreana. Soltera.	¿Dónde vive actualmente? Domicilio y teléfono.	2
Aquí, en Madrid. Calle Gante, número 67, segundo-A. Treinta y ocho, veinticinco, quince, cero cuatro.	Muy bien. Tome asiento y espere a que la llamen.	2
Muy bien. Tome asiento y espere a que la llamen.	Aquí siempre preguntan el apellido materno, ¿te fijas?	2
Me preguntó que dónde y cuándo había nacido y mi estado civil. También me preguntó que dónde vivía.	Parecía enfadado.	2
Ahí está el letrero, es por el pasillo de la izquierda.	¿Qué bueno que haya muchas ventanillas atendiendo, así pasaremos pronto!	2
¿Qué bueno que haya muchas ventanillas atendiendo, así pasaremos pronto!	Es mejor tener el pasaporte listo junto con el formulario que llenamos en el avión, ¿lo tienes ya?	2
No encuentro el formulario, ¿dónde lo puse? ¡Ah! Aquí estaba. Pensé que lo había puesto en mi cartera.	Sus pasaportes, por favor. ¿Son ciudadanas de la Unión Europea?	2
Sí, señor. Somos españolas.	¿Cuál es el objetivo de su viaje?	2
Muy bien. Todo está en orden. Feliz estancia y que disfruten en nuestro país.	Quisiera pasar al servicio, ¿en dónde nos vemos?	2
Donde tú digas.	¿Qué te parece en las bandas para recoger el equipaje?	2
Son estas. Y esas son las de mi amiga.	Pongan aquí su equipaje. ¿Cuántas maletas traen en total?	2
Cuatro, dos grandes, dos pequeñas y nuestras mochilas.	Denme por favor sus declaraciones. ¿Traen algo en especial?	2
No, señor. Nada.	¿De dónde vienen?	2
De España.	Abra esta maleta grande, por favor.	2
Sí, como no. Ana, ayúdame ¿sí? Detenme mi mochila.	¿Llevan tabaco o licores?	2
Solo son botellas de sangría, y unos paquetes de cigarros.	Está bien. Pueden pasar.	2
Está bien. Pueden pasar.	No puedo cerrarla. Ana, intenta tú ¿sí? Me puse un poco nerviosa.	2
No puedo cerrarla. Ana, intenta tú ¿sí? Me puse un poco nerviosa.	A ver, yo le ayudo. Ya está. Pasen, por favor.	2
¿Sí, dígame?	No ha llegado una de mis maletas.	2
No ha llegado una de mis maletas.	¿Revisó bien en la banda giratoria?	2
¿Revisó bien en la banda giratoria?	Sí, señor. Y he esperado hasta el final.	2
Sí, señor. Y he esperado hasta el final.	¿En qué vuelo ha llegado usted?	2
En el vuelo CA219, proveniente de Pekín.	¿Lleva algo de valor en su maleta?	2
No, señor. Solo ropa y objetos personales.	Describame su maleta, por favor.	2
Es rectangular, de color negro. Tiene ruedas y viene cerrada con un candado pequeño. Le he puesto dos listones para distinguirla, uno rojo y otro azul.	Bien. Un momento ... Aquí clave 34. Rectangular, negra con ruedas y candado. Listones azul y rojo. Cambio. Afinitivo. Se encuentra conmigo. Cambio. Esperaremos. Cambio y fuera.	2
Hemos encontrado una que tal vez sea la suya. Aquí viene.	¿He pasado un gran susto! Ahí traía mis documentos para inscribirme en la universidad.	2
¿He pasado un gran susto! Ahí traía mis documentos para inscribirme en la universidad.	¿No me ha dicho usted que no traía nada de valor?	2
Mejor pídeme dos itinerarios, uno para mí también.	¿Billetes de ida y vuelta o sencillos?	2
De ida y vuelta, por favor.	Aquí están. Además de pagar los billetes, deberán cubrir el seguro de viaje.	2
Aquí están. Además de pagar los billetes, deberán cubrir el seguro de viaje.	¿En qué andén sale este tren?	2
En el número seis. Quince minutos antes de la salida deben esperar allí.	Muchas gracias, señor.	2
Muchas gracias, señor.	¿Caramba! Tras cobrar tan caros los billetes todavía tuvimos que pagarle un seguro.	2
Buenos días.	Buenos días. Me permite su boleto, por favor. ¿Prefiere un asiento en ventanilla o en pasillo?	2
Un asiento en ventanilla, en la sección de no fumar, por favor.	¿Va a documentar algún equipaje?	2
Sí, tengo tres maletas y un maletín.	Lo siento, solo puede documentar dos maletas que no excedan los 30 kilos. El exceso de equipaje se cobra extra.	2
Entendido. Entonces, solo documente las dos maletas grandes. ¿Puedo llevar la pequeña y el maletín en la mano?	¿Cuáles son?	2
Gracias. Muy amable.	¿Señorita, olvida algo!	2

Table A33: Examples of labeled data: Textbook (33)

¿Sí, dígame?	Aquí hay un comprobante ...	2
Gracias, Paco. Pasa, por favor.	¡Feliz cumpleaños! Tengo un regalo para ti.	2
¿Te gusta?	¡Me encantan los muñecos de peluche! Muchas gracias, Paco.	2
¡Me encantan los muñecos de peluche! Muchas gracias, Paco.	¿Soy el primero en llegar? ¿Dónde están todos?	2
Bertha Suárez. Mucho gusto.	Los dejo un momento. Voy a la cocina con los muchachos que están preparando unos canapés.	2
Los dejo un momento. Voy a la cocina con los muchachos que están preparando unos canapés.	Me puedes llamar Paco, mis amigos me dicen Paco. ¿Vamos a tomar una copa?	2
Dura unas quince horas.	¿Van a pasar alguna película?	2
Pescado, por favor.	¿Qué le sirvo de tomar?	2
Sí, así es, con vista al mar si es posible.	¿Prefiere una cama matrimonial o dos individuales?	2
No, señor. Este precio solo incluye el desayuno, pero no considera los servicios ni los impuestos.	¿Acepta tarjetas de crédito y cheques de viajero?	2
Sí, señor.	Bueno, nos quedaremos aquí.	2
Bueno, nos quedaremos aquí.	Por favor, llene este formulario y firme el registro. Aquí tiene la llave. El mozo les llevará su equipaje.	2
Por favor, llene este formulario y firme el registro. Aquí tiene la llave. El mozo les llevará su equipaje.	¿Podría enviar una botella de champán a la habitación?	2
Mire, estas son las habitaciones. Les entra mucho sol por la mañana.	¿Cuántos dormitorios hay?	2
Naturalmente.	¿Cuánto cuesta el alquiler?	2
Como modelo compacto de cuatro puertas tenemos el Opel Corsa y el Century Clio. Los dos tienen cuatro asientos y aire acondicionado.	¿Son automáticos?	2
Los tenemos con transmisión automática y también manual.	¿Cuál es la tarifa?, para ver cuál nos conviene ... ¿Cobran por kilometraje?	2
Me quedo con el Opel Corsa automático.	¿Le incluyo el seguro en la factura?	2
Sí, claro, incluya el seguro contra todo riesgo.	Pennitame su licencia de conducir.	2
No, señor. Su licencia local para conducir es suficiente.	Disculpe, señorita, aquí en España cuando se saludan se dan dos besos, ¿verdad?	2
De Estela, su compañera de trabajo.	¿Es algo urgente? ¿Desea dejar algún recado?	2
Sí, por favor. Se trata de una reunión de la facultad. Necesito avisarle de un cambio antes de que haga otro compromiso.	Espere un momento. Voy por el bolígrafo para anotar. Ahora sí.	2
Espere un momento. Voy por el bolígrafo para anotar. Ahora sí.	Por favor, le dice que en vez de ser la reunión la semana próxima, será	2
¿Sí? Pues, no, no lo sabía.	¿Quisieras venir?	2
No tengo nada que hacer mañana... Me encantaría ir.	¡Ah! También tenía que avisarle a Lorena... Le llamaré ahora mismo...	2
Sí, lo es. Estaba en oferta cuando lo compré.	¡Genial!	2
Sí, pero me daba un poco de problemas la batería.	Te felicito por el estreno. Entonces, nos veremos mañana. Chao.	2
Un momento, por favor ...	Muchísimas gracias. Otra cosa.	2
Marque el 07 y después 82, enseguida marque el código de la ciudad.	¿Y para llamar a cobro revertido?	2
Dígame el número y su nombre, por favor.	La clave de Seúl es el 2, el número de casa es 387-1780 y mi nombre es Kim Yunjo.	2
La clave de Seúl es el 2, el número de casa es 387-1780 y mi nombre es Kim Yunjo.	¿Con quién desea hablar?	2
Con quien conteste.	Muy bien, cuelgue y espere un momento.	2
Estoy viendo los vestidos.	¿Qué modelo necesita? ¿Cómo lo prefiere?	2
Pues ... uno de color luminoso y alegre, pero un poco menos brillante que ese azul, y que también sea de seda.	¿Qué talla le doy?	2
La treinta y seis.	Mmmm ... Mire, tenemos estos modelos. ¿Le gusta alguno?	2
No, lo siento.	A ver, me lo voy a probar ... Sí, lo siento bien.	2
Del número 24.	¿Qué le parecen estos?	2
Algunas están saliendo muy malas, pero le voy a dar de las más buenas.	Pues, deme dos kilos de patatas, medio kilo de cebollas y esta otra verdura.	2
A 5 euros.	¿Qué caro! ¿No tiene otra fruta más barata?	2
¿Qué caro! ¿No tiene otra fruta más barata?	Pues ... los melones. A 4 euros. Es la fruta de temporada y es tan muy baratos ahora.	2

Table A34: Examples of labeled data: Textbook (34)

Pues ... los melones. A 4 euros. Es la fruta de temporada y están muy baratos ahora.	Huelen muy bien. Deme uno.	2
Huelen muy bien. Deme uno.	¿Algo más?	2
¿Cuánto es en total?	Son 23 con 34. Y le voy a regalar estas calabacitas. Están bien frescas.	2
Son 23 con 34. Y le voy a regalar estas calabacitas. Están bien frescas.	Aquí tiene, 35 euros. Y los 10 de la cuenta pendiente de la otra vez.	2
Sí, aquí está. A las 9, para dos personas.	Me gustaría cambiarla para las 9 y media, y 4 personas en vez de 2.	2
¿Me permiten sus abrigos?	¿Qué lugar tan exótico! ¿No será demasiado caro?	2
No, no es tan caro. Espero que les guste.	Aquí tienen la carta y la lista de vinos. ¿Les apetece un aperitivo?	2
Sí, tomaremos vermut y jamón serrano.	Yo no bebo. Para mí, una sangría, si no fuera mucha molestia.	2
Y que vayan preparando el especial de merluza a la romana, para 4 personas, pero que no quede salada.	Hoy está frita con una salsa muy suave y setas con nata. Les gustará.	2
Una cerveza para mí y un plato de patatas fritas para acompañar.	Yo quiero una copa de vino tinto.	2
Yo quiero una copa de vino tinto.	¿Algo más?	2
Para mí, otro igual pero con mucha lechuga.	Estamos recibiendo ahora un nuevo tipo de jamón de york sabroso.	2
Yo lo compro los domingos. No me importan tanto las noticias, sino los suplementos que son muy buenos. Pero entre semana suelo comprar otros.	¿Y cómo puedes leer cada día un periódico diferente?	2
¿Y cómo puedes leer cada día un periódico diferente?	En realidad, no me interesa tanto la tendencia política del periódico, sino las secciones en las que está especializado.	2
En realidad, no me interesa tanto la tendencia política del periódico, sino las secciones en las que está especializado.	¿Y por qué no los lees a través de Internet? Así podrías ver lo que más te guste de cada uno de ellos diariamente. ¿no?	2
Es verdad. Pues vamos a ver lo que hay en el periódico de hoy.	Oye, echa un vistazo al pronóstico meteorológico, a ver qué tiempo o vamos a tener este fin de semana.	2
Oye, echa un vistazo al pronóstico meteorológico, a ver qué tiempo vamos a tener este fin de semana.	Toma, te paso la sección.	2
Lo mismo está en un embotellamiento. Ya sabes ... a esta hora hay mucho tráfico.	Si pasáramos por ella, no tendríamos que esperar.	2
Anoche sali y regresé a casa por la madrugada. Y esta mañana estaba tan profundamente dormida que no he podido oír el despertador. Para colmo, el tráfico hoy estaba terrible.	Deberías haberme llamado ... Llevamos una hora esperándote.	2
Antes de salir de casa te llamé, pero tu teléfono estaba comunicando. Lo siento de verdad.	Estaría llamándote a ti. Pero ¿por qué no llamaste después? Tienes celular, ¿no?	2
Con las prisas, lo dejé en casa. Ya sabes, siempre estoy despistada. Lo siento de veras. Esto no volverá a suceder.	Ángela quería ir a buscarte. Si hubiéramos pasado por ti, no habríamos esperado tanto, como dijo Ángela. Pero, bueno, ya pasó todo. Ya es tarde para ir al cine, vámonos a cenar, yo las invito.	2
Y ¿cómo quedamos, entonces?	La última función empieza a las nueve y media, pero si te parece, nos vemos a las siete para cenar.	2
La última función empieza a las nueve y media, pero si te parece, nos vemos a las siete para cenar.	La verdad ... es que no me va bien tan pronto. Si hubiera sabido antes, habría aceptado con mucho gusto.	2
Nos vemos enfrente de "La copa de leche". Alcanzaremos a tomar un café.	A propósito, tendremos que comprar las entradas en la taquilla, ¿no?	2
¿Recuerdas qué bien actuó de malo en la película que vimos el otro día?	Si me hubieras dicho antes, habríamos invitado a Florencia, ya sabes que es fanática de Banderas.	2
Sí, claro. Pasa. De hecho, me iba a tomar uno ahora mismo.	Gracias.	2
Sí, lo vi. Pero fue algo superficial, a tal punto que no explicaba nada sobre los problemas ambientales ni tampoco proponía soluciones. Tan solo quería impresionar a los espectadores.	Entonces, ¿qué tipo de programas te interesan?	2
La verdad, veo un poco de todo porque puedo ver la televisión por cable. También estoy suscrita al Canal Plus.	Según la Guía del ocio, hoy a las diez de la noche ponen una película divertida en el dos. Mira ... está a punto de empezar.	2
No tiene caso, de todos modos van a cortar cualquier programa bueno que encontremos.	Toma, un pedazo del pastel que te gusta. Para que estés contenta ...	2
Está delicioso.	¿Sabías que mañana pasan el partido de fútbol entre Corea y Grecia en el Canal Plus?	2
O sea, que has dejado vacía la tienda ...	Tú no te preocupes.	2
Yo soy el experto, a tal punto que siempre me toca preparar los bocadillos en casa de mis padres.	Bueno.	2
Corea está ganando uno a cero.	¡Espera ... ! ¡Gooooool! Grecia ha metido un gol. Ahora los dos equipos están empatados. ¡Ay, no! ¡Qué mala suerte!	2
¡Es Pak, lanza y ... gooooo! ¡Era imparable!	¿Adónde quieres ir a cenar?	2
Tenga esta forma, llénala con sus datos personales;	Aquí está.	2
Cuéntame cómo fueron los hechos.	¿Algún objeto desaparecido?	2
Los pasaportes, 1.500 euros en cheques de viajero, una tarjeta de crédito y ropa.	¿Tiene seguro de viaje?	2
¿Sí? Entonces, haré como me sugiere y abriré una cuenta en dólares.	¿Quiere una cuenta corriente o de ahorros?	2
Claro que sí, y no le cobramos comisión, como le dije antes.	Por fin, mi esposo estará feliz con la cuenta en dólares, como él quería.	2
Por fin, mi esposo estará feliz con la cuenta en dólares, como él quería.	Aquí tiene su talonario y la tarjeta para el cajero automático.	2

Table A35: Examples of labeled data: Textbook (35)

Sí, le agradezco.	¿Funciona ahora?	2
Hay un taller medio kilómetro más adelante.	Es una buena idea.	2
Espere un momento, lo veré en seguida.	¿Cree que será muy difícil?	2
¿Cree que será muy difícil?	Hace un ruido extraño.	2
Normalmente, el tráfico está menos congestionado en septiembre que en julio y a gusto.	¿No crees que sea mejor tomar la autopista que la carretera local para ir a los Pirineos?	2
Creo que yo sí cumplo con todos los requisitos. Domino bien el inglés y el francés.	¿Tiene usted disponibilidad para viajar?	2
Sí, absolutamente.	Quien puede viajar tiene buenos puntos en la selección.	2
Gracias por su tiempo, hasta mañana.	Los que están afuera han presentado ya esta entrevista.	2
Vamos a ver. .. Abra la boca, por favor.	Como usted crea conveniente, doctor.	2
Le voy a poner un poco de anestesia.	Ahora dígame, ¿cómo se siente?	2
Me caí de una escalera.	¿Cuánto tiempo hace que le pusieron una inyección antitetánica?	2
No recuerdo exactamente.	Entonces, pregunte a quien le puso la vacuna. Luego me avisa.	2
No he podido ir a trabajar. El asistente que me ayuda se ha encargado de todo.	¿Hay alguna persona diabética en su familia?	2
El pescado que comí ayer estaba muy fresco.	Vuelva mañana. Necesito hacerle unos análisis que complementen el diagnóstico.	2

A.5.2 Twitter corpus

Table A36: Examples of labeled data: Twitter (1)

RT @mauriciomacri: No hay discusión. Messi deslumbra siempre. Aplausos para un jugador de otro mundo 🌟	@mauriciomacri Así era! UNICO EN LAS GALAXIES	1
RT @soynapringada_ @Desahogada La Rispa cada vez me da mas vergüenza ajena. Paso que da, mas pena me da. Ni que fuera... https://t.co/3FobYbltWj	@soynapringada_ @Desahogada La Rispa cada vez me da mas vergüenza ajena. Paso que da, mas pena me da. Ni que fuera... https://t.co/3FobYbltWj	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas #AllwaysPcComponentes ¡Ojalá esta preciosidad en mi habitación 🙏	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas #AllwaysPcComponentes Me caigo muertísima si consigo este pedazo sorteo. 🙏	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas increíble el pc	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas Me da un infarto si me llevo esa máquina!!	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas Lo necesito urgentemente 🙏	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas A ver si hay suerte 🙏	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas Ya toca que sea yo el ganador. Aunque no lo sea	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas este es mío!!	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas Ya es hora de tener un PC gamer y cambiar un poco el contenido del canal	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas ya le estoy haciendo un hueco en mi cuarto🙏	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas lo necesito ya	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Lévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/3FobYbltWj	@pccomponentes @elchapuzas Suerte	1
RT @ultrazoquet: El infotunio se apoderó del Paraguay he'i kuri voi Roa Bastos correlli	@ultrazoquet Esa frase es de Teodosio González!! Quien fue ministro y autor del Código Penal anterior correlli	1
RT @ultrazoquet: El infotunio se apoderó del Paraguay he'i kuri voi Roa Bastos correlli	@ultrazoquet Así es!! Infotunio!	1
RT @ultrazoquet: El infotunio se apoderó del Paraguay he'i kuri voi Roa Bastos correlli	@ultrazoquet Correlli, la mufa de cierto jefe (nos prohíben ahora hablar mal de...)	1
RT @Lhuchin: Chileno, no te olvides que el CSM que sacó a Bielsa fue Sebastián Piñerá	@Lhuchin @Sergio_Digital Y eso que tiene que ver si al final: él que ganó algo fue Saipaoli y Bielsa se fue solito.	1
RT @PabloMM: Es alucinante que Albert Rivera se auto proclame portavoz de la mayoría de los catalanes cuando tiene en Cataluña los mismos A...	@PabloMM Y que yo.	1
RT @PabloMM: Es alucinante que Albert Rivera se auto proclame portavoz de la mayoría de los catalanes cuando tiene en Cataluña los mismos A...	@PabloMM @gorkaotxo Nos ha recordado otra vez que es catalán...?	1
RT @Oficialmorris: Un PRI que se multiplicó y tuvo muchos mini PRIs https://t.co/3r8aFbkTbp	@Oficialmorris Esa foto tuya es vieja! ya estas mas grande no te hagas	1
RT @Oficialmorris: Un PRI que se multiplicó y tuvo muchos mini PRIs https://t.co/3r8aFbkTbp	@Oficialmorris Y el PES aún no cuenta?	1
RT @danielscioli: AR/IVAMOS ARGENTINA DERECHO A RUSIA!ru https://t.co/sIA9k1tj11	@danielscioli Seguro que te le diste un amNo...ChorroVos sos cualquier cosa. Menos Argentino.	1
RT @xcrstn: si habéis follado con una piba y os corréis pero ella no se corre no os la 'habéis follado', os ha follado ella a vosotros gili...	@xcrstn pues entonces a mi me han follado más veces a de las que he follado yo...	1
RT @maiaparker_ @rubelangelifturu Nos odias xq no.podes ser como nosotros :)	@maiaparker_ @rubelangelifturu Tu sabes	1
RT @ElRojoEstaGede: 3 cosas que deben morirse ya mismo:- racing- chile- el periodismo	@ElRojoEstaGede pobre. racing dejalo que es mi puta loca.	1
RT @rosadiezglez: Extracto del editorial sobre Cataluña del semanario satirico frances 'Charlie-Hebdo'. https://t.co/aTY89U54VF	@rosadiezglez @XupxoGN pues por el último párrafo está muy claro y lo vio todo el mundo	1
RT @rosadiezglez: Extracto del editorial sobre Cataluña del semanario satirico frances 'Charlie-Hebdo'. https://t.co/aTY89U54VF	@rosadiezglez Muy cierto. 🙏	1
RT @rosadiezglez: Extracto del editorial sobre Cataluña del semanario satirico frances 'Charlie-Hebdo'. https://t.co/aTY89U54VF	@rosadiezglez Es tan obvio, tan obvio, que poca gente sabe verlo. Y menos los que tienen intereses en romper España... https://t.co/SKPxsWTFX5	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no son mas que puro teatro.después se sientan a tomar whiskey y a reír...	@PirryTv Pero obvio, todo es un show	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no son mas que puro teatro.después se sientan a tomar whiskey y a reír...	@PirryTv Exactamente eso era lo que decía Rojas Pinilla, que sabía perfectamente como funcionaba el circo.	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no son mas que puro teatro.después se sientan a tomar whiskey y a reír...	@PirryTv Igual pienso yo...	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no son mas que puro teatro.después se sientan a tomar whiskey y a reír...	@PirryTv Pense que era la única que pensaba eso ...	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no son mas que puro teatro.después se sientan a tomar whiskey y a reír...	@PirryTv Aun lo dudas. Esa es la politica tradicional embarracar al pueblo ofrecerle un tamal y listo votan por el candidato que sea. Triste realidad.	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no son mas que puro teatro.después se sientan a tomar whiskey y a reír...	@PirryTv De acuerdo	1
RT @LadyGhostVzla: Siempre se hacen bromas de la mala memoria del venezolano. pero jamás pensé que muchos se olvidarían de la lucha de ello...	@LadyGhostVzla 🙏🙏🙏🙏🙏🙏 Yo no olvidaré!!!!	1
RT @Elaguion_ Y esto es lo que piensa Pablo Iglesias sobre España, sin guiño, sin cámaras. sin montaje. https://t.co/RudGz8jhTj	@Elaguion_ Pero ahí le tenemos...es lo increíble, 5 millones de votos le avalan, cosa que a mi me tiene estupefacto	1
RT @Elaguion_ Y esto es lo que piensa Pablo Iglesias sobre España, sin guiño, sin cámaras. sin montaje. https://t.co/RudGz8jhTj	@Elaguion_ Sin comentarios	1
RT @SOMEXICAN: Éste wey era yo en todas las fiestas familiares. https://t.co/2Wlhfq3gd	@SOMEXICAN Ponian son sillas juntas y ahí te acostaban @MaciasMaris #goodtimes	1
RT @numer344: Si cometo un robo y me pilla la policía, puedo suspender el robo para dialogar con la policía?	@numer344 Sólo si el intento de robo no sobrepasa los 8 segundos	1

Table A37: Examples of labeled data: Twitter (2)

RT @numer344: Si cometo un robo y me pilla la policía, puedo suspender el robo para dialogar con la policía?	@numer344 @forogc ¡Deberías!	1
RT @numer344: Si cometo un robo y me pilla la policía, puedo suspender el robo para dialogar con la policía?	@numer344 @forogc Tu dices que te has liado con el robo si y robo no, y la lías t e parda (sin querer).	1
RT @numer344: Si cometo un robo y me pilla la policía, puedo suspender el robo para dialogar con la policía?	@numer344 @roxiero72 Es que es surrealista, la verdad.	1
RT @numer344: Si cometo un robo y me pilla la policía, puedo suspender el robo para dialogar con la policía?	@numer344 Demasiadas facilidades le dan	1
@ChuchoRZ Ya te van a decir que no existe la clase media, missing the whole poi nt del tema que estas trayendo a colación. 🙄🙄🙄🙄	@ChuchoRZ No entiendo por qué pagar por algo que puedes recibir gratis. Eso es lo insólito.	1
RT @PabloMM: Es alucinante que Albert Rivera se auto proclame portavoz de la m ayoría de los catalanes cuando tiene en Cataluña los mismos A...	@PabloMM Es que esta noche ha soñado y aún no se ha despejado.	1
RT @PabloMM: Es alucinante que Albert Rivera se auto proclame portavoz de la m ayoría de los catalanes cuando tiene en Cataluña los mismos A...	@PabloMM Rivera nunca será querido por los catalanes, y mucho menos por los e spañoles, ese será su castigo político.	1
Hermanos de Panamá y Honduras. @RusoEl23 lo dijo hace semanas y se cumplió. ¡Al Mundial y Repechaje!	@FOXSportsMX @RusoEl23 @UpalabraMX @RusoEl23 A veces haces uno coment arios que molestan pero este día la sacaste del estadio 🙄	1
Hermanos de Panamá y Honduras. @RusoEl23 lo dijo hace semanas y se cumplió. ¡Al Mundial y Repechaje!	@FOXSportsMX @RusoEl23 @UpalabraMX Gracias excelente pronóstico. Panamá a l mundial.	1
RT @martindelp: Si eres aficionado a las Chivas y te parece que Lainez es malísimo por ser del América, no te gusta el fútbol, sólo las Chi...	@martindelp No es malísimo, pero no es la súper estrella.Ojalá y sea constante pe ro están tan necesitados de un crack...	1
RT @martindelp: Si eres aficionado a las Chivas y te parece que Lainez es malísimo por ser del América, no te gusta el fútbol, sólo las Chi...	@martindelp Juega en el América lo van a criticar juegue donde sea. De este nivel de odio genera el mas grande (América)	1
RT @yosoyনারanjito: El Jueves y Charlie Hebo dos formas de afrontar el problem a nacionalista. Uno aprovecha la ocasión para aremeter con...	@yosoyনারanjito. Volvera el jueves cuando pase hambre, todo a su tiempo	1
RT @yosoyনারanjito: El Jueves y Charlie Hebo dos formas de afrontar el problem a nacionalista. Uno aprovecha la ocasión para aremeter con...	@yosoyনারanjito. Se acabo El Jueves, una cosa es la sátira y otra es pintar siempre a España como el origen de los problemas de los demas.	1
RT @tinocotania: Como te hemos recordado Bolillo. Con gratitud y aplaudiendo la hazaña panameña. https://t.co/3XQ5Erbgr	@tinocotania Felicitaciones Bolillo 🇸🇨🇸🇨🇸🇸🇨	1
RT @publico_es: Lluvia de multas de 601 euros a vecinos que se sentaron en las vi as para protestar contra el 'muro' de Murcia https://t.co/...	@publico_es @FjRamí21 ya verás los fachas de Valencia, que se vayan preparand o 🇪🇸🇪🇸🇪🇸	1
@CARLOSLGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos p or la revancha en Rusia!! ... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSLGUERRERO Vamos por la revancha en Rusia... aaah... changos... bueno, nos vemos en 2022!! 🇺🇸🇺🇸	1
@CARLOSLGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos p or la revancha en Rusia!! ... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSLGUERRERO Rusia no los quiere!!	1
@CARLOSLGUERRERO Y que dices de Alemania, ellos siguen vivos y Mejor que nu ca, no es karma simplemente es dejar de... https://t.co/VWGN84ffaa	@CARLOSLGUERRERO Los miserables! 🇩🇪	1
RT @tomy_0k: -Sali payaso puto... https://t.co/55q8c5XzQG	@tomy_0k. Ayer la vill!! Me hiciste reír!! 🇪🇸🇪🇸🇪🇸🇪🇸	1
RT @LaHipogriña: Albert Rivera: "El nacionalismo divide a la gente". Pero sólo el ca talán, el español no. Ese nada. #11Oct.	@LaHipogriña Que razonamiento tan estúpido.	1
RT @LaHipogriña: Albert Rivera: "El nacionalismo divide a la gente". Pero sólo el ca talán, el español no. Ese nada. #11Oct.	@LaHipogriña Hahaha 🇪🇸🇪🇸🇪🇸🇪🇸🇪🇸 Siempre quise decir eso pero en 140 cará cteres no sabía cómo... gracias. 🇺🇸	1
@ChuchoRZ Ya te van a decir que no existe la clase media, missing the whole poi nt del tema que estas trayendo a colación. 🙄🙄🙄🙄	@ChuchoRZ No entiendo por qué pagar por algo que puedes recibir gratis. Eso es lo insólito.	1
RT @NoblexAr: Aclaremos desde ahora: La #PromoNoblex no era valida para Chile 🇨🇱 #Rusia2018 https://t.co/zMKxC4La78	@NoblexAr Señor Noblex, usted es diabólico https://t.co/RtrJAONwNr	1
RT @PadreJosePalmar: Este 15-Oct votar temprano garantizará que el proceso elec toral no pueda ser sabotado por los narcos de Miraflores. N...	@PadreJosePalmar Así es, votaré en contra de la dictadura. No por un nombre o u na cara. Ellos quieren que no salgamos... https://t.co/dy5Yw6eORk	1
RT @Edy5mol: Aquí con la más guapa del mundo 🇲🇽 @LuceroMexico en exclusiva para todos ustedes 🇲🇽🇲🇽🇲🇽🇲🇽🇲🇽 la más cool del mundo !!🇲🇽🇲🇽🇲🇽🇲🇽🇲🇽 https://t.c...	@Edy5mol @LuceroMexico Hermosa mujer, esa sonrisa ilumina el día de todo mu ndo!	1
RT @gabrielrufian: Entiendo que si Rajoy dice que el referéndum se desarticuló el 30/09, el 1/10 se apaleó a la gente por puro sadismo.	@gabrielrufian nadie hace nada	1
RT @GianThurber: Los ticos son los mejores vecinos que la vida nos pudo dar, cele braron y alentaron a la sele como un panameño más.	@GianThurber Es muy cierto porque nos visitan y no se quedan como remoras op ortunistas de nuestra solidaridad.	1
RT @Diszain: #MasterChefMx no puedo creer que ganó el equipo rojo 🇲🇽, tendre mos que soportar a doña Julia una semana más 🇲🇽 https://t.co/UCW6...	@Diszain El Padre Beto fue lo mejor, jajaja.	1
RT @Diszain: #MasterChefMx no puedo creer que ganó el equipo rojo 🇲🇽, tendre mos que soportar a doña Julia una semana más 🇲🇽 https://t.co/UCW6...	@Diszain Te la bañaste!!! Jajajajaja	1
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de escritorio, Co lombia y Perú en el min88 tenían que jugar a clasificar a...	@fdbedout Y no tienes nada más que hacer? Pasa la pagina	1
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de escritorio, Co lombia y Perú en el min88 tenían que jugar a clasificar a...	@fdbedout Cada uno lo ve según sus conveniencias... cuando nos sirve esta bien p ero cuando nos afecta el análisis es diferente	1
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de escritorio, Co lombia y Perú en el min88 tenían que jugar a clasificar a...	@fdbedout No hay pruebas de acuerdo en Lima, algun audio? en derecho nadie s abe que hablan los jugadores, los lectores de labios lo dicen.	1
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de escritorio, Co lombia y Perú en el min88 tenían que jugar a clasificar a...	@fdbedout Si están locos los chilenos que piensen eso. De malas para que no jug aron bien. Y ahora con todos esos borracos in disciplinados que más querian.	1
RT @EmiRigoni: Muy feliz por la clasificación !!El mejor de todos los tiemposLe gu ste a quien le guste 🇨🇱🇲🇽 https://t.co/rMjedUwXZl	@EmiRigoni @JulitaArgento. El mejor jugador de toda la historia, que pedazo de mostro, vos crack papa, felicitaciones y gracias por darnos esta alegría.	1
RT @EmiRigoni: Muy feliz por la clasificación !!El mejor de todos los tiemposLe gu ste a quien le guste 🇨🇱🇲🇽 https://t.co/rMjedUwXZl	@EmiRigoni Los amo a los dos Ertoy super feliz porque clasificaron y les dijeron a todo el mundo que nosotros somos los mejores!!	1
RT @dlacalle: Mi intervención esta mañana desde Londres en @CNBC desmontand o los argumentos independentistas (!) https://t.co/XnhLT338K9	@dlacalle @jaimeberenguer @CNBC Enhorabuena por defender España y por ese magnifico ingles	1
RT @SusodeToro1: siguen cerrando webs, grupos de whatsapp, el objetivo final es la radio y televisión públicas catalanas. Sintonicemos mien...	@SusodeToro1 Ellos rompen las leyes y los derechos más basicos de la democraci a pero las cargas policiales son para el pueblo.	1
RT @Vitalisioo: RT si la ves gris y celesteMG si la ves rosa y blanca https://t.co/BHF eAsfVgV	@Vitalisioo @agustinreyes14 Ay, ahora la veo gris y celeste @agustinreyes14 @ffi arenciamonti @Rojas_Agos	1
RT @Vitalisioo: RT si la ves gris y celesteMG si la ves rosa y blanca https://t.co/BHF eAsfVgV	@Vitalisioo quiero cambiar mi voto, ahora la veo rosa y blanco	1
RT @MarcosWitt: "No dejen que el corazón se les llene de angustia; confíen en Di os y confíen también en mí (Jesús)." Juan 14:1	@MarcosWitt Amén	1

Table A38: Examples of labeled data: Twitter (3)

@Wewerwero Más que karma es porque ellos se parten el lomo contra B razil, Italia, Colombia, Argentina, etc. Y Mex... https://t.co/UboAaUmcBz	@Wewerwero A ver pinche cabra, chile por que? De Holanda te creo... au n que la desepcion mejicana son troncos hasta una piedra juega mejor.	1
@Wewerwero Más que karma es porque ellos se parten el lomo contra B razil, Italia, Colombia, Argentina, etc. Y Mex... https://t.co/UboAaUmcBz	@Wewerwero Chile karma de que? En su momento ellos le ganaron a la selección mexicana de forma justa	1
@Wewerwero Más que karma es porque ellos se parten el lomo contra B razil, Italia, Colombia, Argentina, etc. Y Mex... https://t.co/UboAaUmcBz	@Wewerwero Es cierto,👍	1
RT @Vitalisio: RT si la ves gris y celesteMG si la ves rosa y blanca https://t.co/8HFeAsfWgv	@Vitalisio Yo veo un vestido blanco y dorado	1
RT @Vitalisio: RT si la ves gris y celesteMG si la ves rosa y blanca https://t.co/8HFeAsfWgv	@Vitalisio Hace un momento la he visto azul y ahora rosa como normal mente lol	1
RT @Vitalisio: RT si la ves gris y celesteMG si la ves rosa y blanca https://t.co/8HFeAsfWgv	@Vitalisio Mi mamá dice que es verde, yo ya no sé que creer	1
RT @VicencNavarro: Después de llamarme secesionista, espero que @el_p ais rectifique esta falsedad por dignidad y ética periodística. https://t.co/8HFeAsfWgv	@VicencNavarro @el_pais Vicenc, me temo que de eso no les queda...	1
@Lenny_Caceres El tipo fue un genio desde el minuto 1 en que sacó la pr omo	@Lenny_Caceres Ahora es socio	1
@Lenny_Caceres El tipo fue un genio desde el minuto 1 en que sacó la pr omo	@Lenny_Caceres Jajajaja si totalmente	1
RT @AlexNoker: Tranquilos, no pasará nada, ni habla de Carrero Blanco ni de españoles de bien https://t.co/q3WD3f9nrl	@AlexNoker @PablitoCastill2 Esto para los rastreadores	1
RT @AlexNoker: Tranquilos, no pasará nada, ni habla de Carrero Blanco ni de españoles de bien https://t.co/q3WD3f9nrl	@AlexNoker No será este hombre un genio asociando conceptos	1
RT @AlbertoRT51: ÚLTIMA HORA MUD pedirá a la comunidad internacio nal más sanciones contra rectoras del CNE y sus familiares por violar de...	@AlbertoRT51 Perfecto!	1
RT @Noemi_delaCalle: Que Pablo Iglesias dedique a @Albert_Rivera la pri mera frase da cuenta de quién está poniendo el sentido común en la...	@Noemi_delaCalle @Albert_Rivera Pienso igual!	1
RT @Noemi_delaCalle: Que Pablo Iglesias dedique a @Albert_Rivera la pri mera frase da cuenta de quién está poniendo el sentido común en la...	@Noemi_delaCalle @Albert_Rivera Que pena poner los países por delante de las personas. Rivera es un arribista.	1
RT @emilianodiaz34: Orgullosos de ustedes! Son los que dan la cara y sop ortan infinidad de cosas por solo defender a su país #vamosy grac...	@emilianodiaz34 Eramos felices con los Díaz y no lo sabíamos👍	1
RT @Coculo: Que la alegría de clasificarnos al Mundial no nos haga olvida r que Massa es un pelotudo.	@Coculo Y que si Messi se resfria qué hacemos.	1
RT @AcunaMarcos17: Uno de los días más felices de mi vida! Vamos Arge ntina an https://t.co/h3zCvXHHbl	@AcunaMarcos17 Gracias.. muchas gracias	1
@MarioBautistaMB: Tu sonrisa te dará un semblante positivo que hará qu e la gente se sienta mejor a tu alrededor.	@MarioBautistaMB Ahora entiendo porque me siento tan bien contigo	1
@MarioBautistaMB: Tu sonrisa te dará un semblante positivo que hará qu e la gente se sienta mejor a tu alrededor.	@MarioBautistaMB Como tú	1
RT @intenseix: les juro que no sé si reir o llorar https://t.co/V5kXiwrVGO	@intenseix Ay que bello👍👍👍	1
RT @intenseix: les juro que no sé si reir o llorar https://t.co/V5kXiwrVGO	@intenseix @babymihyxk no sé pero nos llamamos igual ahora	1
RT @intenseix: les juro que no sé si reir o llorar https://t.co/V5kXiwrVGO	@intenseix Jajajaja lo amo	1
RT @Serranoismael: Pero si lo envuelves en una bandera apenas se nota. https://t.co/t575jxx0M1	@Serranoismael @alexsandra_sanv En esto es en lo que se nota la unidad de España...puedes recorrerla entera	1
RT @Serranoismael: Pero si lo envuelves en una bandera apenas se nota. https://t.co/t575jxx0M1	@Serranoismael Se puede criticar al gobierno pero la bandera no es propi edad exclusiva de nadie es de todos.respetenla	1
RT @porlagoma: Se fueron de bruces promoviendo la abstención, pero se dieron cuenta de que eso beneficia al régimen y enmendaron, bienvenid...	@porlagoma Quienes? Yo no voy a votar .	1
RT @BellaParejaLyL: Rico martes para todos. Aquí Mojada de frente y un p or detras, Besitos, ya viene el fin de semana, veremos como se pone...	@BellaParejaLyL le queda muy bien ese traje de baño	1
RT @pollo_va: Disturbios en Zurich tras la derrota contra Portugal. Los in adaptados de siempre. ch 🇮🇹 https://t.co/DGK2xwp3YN	@pollo_va Que salvaje, igual que aquí en España...	1
RT @TomasCivetta: Que partido te perdiste Maldonado https://t.co/dRzxx5os2L	@TomasCivetta Hay que ser pelotudo para gritar eso... y para difundirlo u n pelotudo al cuadrado...	1
RT @esdonoso_: CTM!!!!ba en una 212 y un abuelito tocaba armónica la z orra, al bajarme le di billete de luca y me dijo: te pedi plata acas...	@esdonoso_ Jajaja esa reacción es más común de lo que se cree	1
RT @esdonoso_: CTM!!!!ba en una 212 y un abuelito tocaba armónica la z orra, al bajarme le di billete de luca y me dijo: te pedi plata acas...	@esdonoso_ mas allá de que a lo mejor no lo haya pedido, no es forma d e responder a nadie.. mala onda, tu lo hiciste con buena intención	1
RT @esdonoso_: CTM!!!!ba en una 212 y un abuelito tocaba armónica la z orra, al bajarme le di billete de luca y me dijo: te pedi plata acas...	@esdonoso_ Creo que me hiciste el día	1
RT @esdonoso_: CTM!!!!ba en una 212 y un abuelito tocaba armónica la z orra, al bajarme le di billete de luca y me dijo: te pedi plata acas...	@esdonoso_ @A_NonGratA_ Es como un microcuento, tremendo! Jajajaj	1
RT @Xuxipc: La declaración de Puigdemont ha dejado a todos descolocad os, menos a Rivera.	@Xuxipc Cuando hicieron el casting para el sexto sentido, Rivera quedó e n segundo lugar para hacer del niño. Ahora usa sus poderes para la politi ca.	1
RT @MonsieurSansFoy: El 7-10-34, las tropas de esa II República entraron a tiros en la Generalitat y se llevaron detenido a Companys y a su...	@MonsieurSansFoy Aquí no hay nada más que mentirosos y embaucador es! A ver quien miente más para gatear al poder.	1
RT @MonsieurSansFoy: El 7-10-34, las tropas de esa II República entraron a tiros en la Generalitat y se llevaron detenido a Companys y a su...	@MonsieurSansFoy Que ignorante!!	1
RT @MonsieurSansFoy: El 7-10-34, las tropas de esa II República entraron a tiros en la Generalitat y se llevaron detenido a Companys y a su...	@MonsieurSansFoy Sólo le han faltado caracteres para decirle:Te enteras gilipoyas? Analfabeta y manipuladora?Pues..eal! Ya lo digo yo!!	1
RT @MarinaAlbiol: No falla: llega el debate sobre abusos bancarios y dere cho a la vivienda y los eurodiputados del PP y PSOE desaparecen de...	@MarinaAlbiol @quimsalvador Será que tienen algo que tapar??	1
RT @MarinaAlbiol: No falla: llega el debate sobre abusos bancarios y dere cho a la vivienda y los eurodiputados del PP y PSOE desaparecen de...	@MarinaAlbiol si ya conocemos todos, este tarannà PPSOE, todo lo que s ea enfermarse a la banca! a que si veo! pleitesia política a la banca y no a l pueblo.	1
RT @MarinaAlbiol: No falla: llega el debate sobre abusos bancarios y dere cho a la vivienda y los eurodiputados del PP y PSOE desaparecen de...	@MarinaAlbiol Pero no son los salvadores de España?	1
RT @JoseLChilavert_: @joaquinfalco La de tu esposa mejor, es mas rica	@JoseLChilavert_ @joaquinfalco jajajajaja buenísimo	1

Table A39: Examples of labeled data: Twitter (4)

RT @Irene_Montero_: No hay respuesta más valiente y contundente que escuchar y dialogar. Hay solución: referéndum pactado y con garantías h...	@Irene_Montero_ Uña mensaje que no tiene desperdicio: Si al diálogo!!!	1
RT @Irene_Montero_: No hay respuesta más valiente y contundente que escuchar y dialogar. Hay solución: referéndum pactado y con garantías h...	@Irene_Montero_ Referendum en toda España supongo. Cataluña es sólo de los catalanes? Obvio que no	1
RT @trecet: El OXIMORON PERFECTO: SOMOS INDEPENDENTISTAS SIN FRONTERAS Anna Gabriel	@trecet @sabor_umami En teoría un oximoron te hace pensar en metáforas y por mucho que busco el tercer significado... no lo encuentro	1
RT @CATERINAV: #10Oct URGENTE! Oposición denuncia que CNE acaba de reubicar 205 centros en todo el país, de votación claramente opositora....	@CATERINAV Ojala que lo hayan hecho	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Esas manos ... 😊 Ale te adoro	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Para todo cariño...para todo !!!!🙏🙏🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Hola Ale 🙏 para saludarte!	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Yo las quiero utilizar para dar amor y mas si es hacia ti 🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz 🙏🙏 como el lobo!! Para tocarte mejor!!!! 🙏 en sueños!!	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Acariar, abrazar "y" dar palmas 🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Para aplaudirte 🙏🙏🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Acariar y abrazar, para sentir aún más	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Pues...las tres opciones me parecen estupendas	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Yo amo para lo que las usas tu, para escribirnos bellas letras y regalarnoslas a tus fans que te amamos y para todo lo anterior dicho	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Para todo acariarte y abrazarte poeta en la distancia con mis palabras.Te quiero.Te extraño 🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Te enseño mis manos...	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Te amo🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Yo quisiera tocar las tuyas...♥♥	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Para abrazar!!	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Leerte y derretirse...@LaEsenciaDeSanz	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Ahora comprendemos que tus manos son el camino para hacer llegar tus caricias al alma ♥	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz O hacer magia...♥ https://t.co/kgDHI77aCq	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz @AlejandroSanz puede ser para enviarte un beso? ♥	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Algún día estrecharé tu mano, poeta. Tus canciones, sin saberlo, cuentan mi historia.🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz para ayudar!!🙏🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Has visto que chico tan guapo vi en la Fundación Lo que la verdad importa 🙏	1
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariar, abrazar o dar palmas 🙏 https://t.co/kPw5QN2N8f	@AlejandroSanz Para escribirte cada día lo que te admiro!!!! Y las tuyas me entregan tu alma en forma de canciones increíbles 🙏🙏🙏🙏🙏	1
RT @fedeposik: Me saqué una foto con mi viejo porque cumple años y salió mi vieja atrás re preocupada MIRANDO EL SULTAN https://t.co/6S3eEf...	@fedeposik @whitesheep7w Jajajajaja no se quién eres, pero gracias 🙏	1
RT @CowboyEnParo: Esta es la idea, si. https://t.co/KTANaae8QW	@CowboyEnParo @sykrmela y así se soluciona todo? Amos...	1
RT @DANIROVIRA: ¡Qué gran noticia! 🙏🙏🙏 https://t.co/VjgQDFEGKE	@DANIROVIRA Bien 🙏🙏	1
RT @DANIROVIRA: ¡Qué gran noticia! 🙏🙏🙏 https://t.co/VjgQDFEGKE	@DANIROVIRA ¡Estamos encantados con la noticia! Gracias a todos los que protegéis a los animales 🙏	1
RT @EP_President: 4)Nadie en la Unión Europea va a reconocer esa declaración.	@EP_President Y tu eres...? Nadie	1
RT @agarzon: La derecha española no está asustada. Está mucho más que emocionada: llevaba meses esperando este momento.	@agarzon @ACalvoFer Esta derecha es inconsciente. Ha provocado la situación cortando todos los amigos de solución.	1
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los fliares d l as victimas d once cuando vieron a De Vido preso.Seguiremos luch...	@MariaLujan_Rey Lo primero que pensé es #JusticiaPorLoDeOnce	1
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los fliares d l as victimas d once cuando vieron a De Vido preso.Seguiremos luch...	@MariaLujan_Rey @AlfredoDarrigo Abrazo enorme María Luján, Dios los cuide.	1
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los fliares d l as victimas d once cuando vieron a De Vido preso.Seguiremos luch...	@MariaLujan_Rey Siempre junto a cada uno de Uds Besos	1
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los fliares d l as victimas d once cuando vieron a De Vido preso.Seguiremos luch...	@MariaLujan_Rey @AngeldebritoOk Imposible no hacerlo!!! Ustedes ayudan a que se diera por suerte	1
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los fliares d l as victimas d once cuando vieron a De Vido preso.Seguiremos luch...	@MariaLujan_Rey @AngeldebritoOk que devuelva lo que robo!!!! la vida de esas personas no se puede pero por lo menos que no se quede con la plata	1
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los fliares d l as victimas d once cuando vieron a De Vido preso.Seguiremos luch...	@MariaLujan_Rey Siempre a vuestro lado ¡! Sigamos luchando por tener Justicia ¡! JUSTICIA con todos los corruptos ¡!	1
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los fliares d l as victimas d once cuando vieron a De Vido preso.Seguiremos luch...	@MariaLujan_Rey No pude no recordar cada lágrima derramada por cada uno de los familiares de esas almas... y no pude no REGOCIARME!!! 🙏	1

Table A40: Examples of labeled data: Twitter (5)

RT @nathcampost: Yo ayer después de ver Coco 🐼💖🐼 https://t.co/25cnJKG15g	@nathcampost Que hermosa eres bebita te amo tanto 🐼💖🐼	1
RT @nathcampost: Yo ayer después de ver Coco 🐼💖🐼 https://t.co/25cnJKG15g	@nathcampost Te querías llevar todo, jaja. 🐼👉🐼	1
@JEYAMAMMON: Gracias eternas a todos seguimos!	@JEYAMAMMON Jey te lo mereces me encanta tu forma de ser y tu gran humildad	1
RT @nathcampost: Yo ayer después de ver Coco 🐼💖🐼 https://t.co/25cnJKG15g	@nathcampost JAJAJAJAJAJA Me encantas, eres una bebé	1
RT @lavozdegalecia: 🇸🇰 El Valle de Arán anuncia un pleno el lunes para independizarse de Cataluña https://t.co/nFdi3ndt8 https://t.co/wUTVD...	@lavozdegalecia Voy a pedir la intendencia de mi barrio	1
RT @lavozdegalecia: 🇸🇰 El Valle de Arán anuncia un pleno el lunes para independizarse de Cataluña https://t.co/nFdi3ndt8 https://t.co/wUTVD...	@lavozdegalecia ¿y quieren independizarse de verdad o se van a quedar con el gobierno de corruptos?🇸🇰	1
RT @lavozdegalecia: 🇸🇰 El Valle de Arán anuncia un pleno el lunes para independizarse de Cataluña https://t.co/nFdi3ndt8 https://t.co/wUTVD...	@lavozdegalecia Hay barnos Del Valle de Aran, que no los dejan votar en el referéndum local, aplicarán el 12m. como voten.	1
RT @lavozdegalecia: 🇸🇰 El Valle de Arán anuncia un pleno el lunes para independizarse de Cataluña https://t.co/nFdi3ndt8 https://t.co/wUTVD...	@lavozdegalecia @edubayon, A ver si les dejan hablar aranes allí 🇸🇰	1
RT @lavozdegalecia: 🇸🇰 El Valle de Arán anuncia un pleno el lunes para independizarse de Cataluña https://t.co/nFdi3ndt8 https://t.co/wUTVD...	@lavozdegalecia Que jodido lo van a tener nuestros hijos: imposible aprobar geografía ibérica	1
RT @JesusCintora: La crisis catalana tapa las conclusiones de años de investigación es por corrupción https://t.co/RidkVvWnWc	@JesusCintora Un gobierno atragantado por la corrupción no puede actuar en países democrático y no puede actuar con La concordia que asume el Estado	1
RT @JesusCintora: La crisis catalana tapa las conclusiones de años de investigación es por corrupción https://t.co/RidkVvWnWc	@JesusCintora De eso se trataba no ?.cualquiera diría que todo calculado	1
RT @barceloh25: Los senadores del PP jalean a las medidas del 155 mientras Rajoy y las enumera. 🇸🇰	@barceloh25 Seguro que ya les ha jaleado los Si en la votación de la independencia pero en el baño de tu casa, secreto, como los que votan.	1
RT @barceloh25: Los senadores del PP jalean a las medidas del 155 mientras Rajoy y las enumera. 🇸🇰	@barceloh25 @democraciareal un partido que necesita respaldarse en sus acciones para que el público les crea	1
RT @NomaskMail: Echarrí no se fue del país.Brancatelli tampoco.Victor Hugo menos Hebe no se suicidó.Promesas de campaña incumplidas:	@NomaskMail Ven... pero se perdieron las elecciones..la gente no les cree!!!! Son todos los K: los que nos gobernaron y sus fieles adeptos!	1
RT @NomaskMail: Echarrí no se fue del país.Brancatelli tampoco.Victor Hugo menos Hebe no se suicidó.Promesas de campaña incumplidas:	@NomaskMail Adonde van a ir estos inútiles...solo acá tienen pantalla.Argentina paraísos generoso	1
RT @NomaskMail: Echarrí no se fue del país.Brancatelli tampoco.Victor Hugo menos Hebe no se suicidó.Promesas de campaña incumplidas:	@NomaskMail @carlos91005124 Verdad no cumplidas!!	1
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, muchas gracias por tenerme en el show lo disfruté mucho @latinamas20...	@Camila_Cabello @LatinAMAs Estuviste maravillosa	1
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, muchas gracias por tenerme en el show lo disfruté mucho @latinamas20...	@Camila_Cabello @LatinAMAs Estuviste increíble hermosa!!!💖💖	1
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, muchas gracias por tenerme en el show lo disfruté mucho @latinamas20...	@Camila_Cabello @LatinAMAs TE AMO	1
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, muchas gracias por tenerme en el show lo disfruté mucho @latinamas20...	@Camila_Cabello @LatinAMAs Guapa!! Estoy orgullosa de ti, te quiero!💖💖	1
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, muchas gracias por tenerme en el show lo disfruté mucho @latinamas20...	@Camila_Cabello @LatinAMAs Lo hiciste increíble chica, cada día mejor y cada día te quiero más.	1
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, muchas gracias por tenerme en el show lo disfruté mucho @latinamas20...	@Camila_Cabello @CCabelloNews @LatinAMAs Lo hiciste increíble💖💖💖	1
RT @davidc_01: Un abrazo fraternal des de la República Catalana a todos los pueblos de España. Hemos roto el candado del 78. con y por vos...	@davidc_01 yo tambien soy pueblo de españa? no noto haber ganado nada con vuestra estultia republica	1
RT @davidc_01: Un abrazo fraternal des de la República Catalana a todos los pueblos de España. Hemos roto el candado del 78. con y por vos...	@davidc_01 Joder, pues ya que os ponéis a romper candados podríais también a brrr libros, digo yo...	1
RT @davidc_01: Un abrazo fraternal des de la República Catalana a todos los pueblos de España. Hemos roto el candado del 78. con y por vos...	@davidc_01 Republica que???? Vamos me lo cuentas mañana.	1
RT @davidc_01: Un abrazo fraternal des de la República Catalana a todos los pueblos de España. Hemos roto el candado del 78. con y por vos...	@davidc_01 Suerte!	1
RT @davidc_01: Un abrazo fraternal des de la República Catalana a todos los pueblos de España. Hemos roto el candado del 78. con y por vos...	@davidc_01 que candado el del armano? así os va.	1
RT @davidc_01: Un abrazo fraternal des de la República Catalana a todos los pueblos de España. Hemos roto el candado del 78. con y por vos...	@davidc_01 Viva La República !!!!	1
RT @davidc_01: Un abrazo fraternal des de la República Catalana a todos los pueblos de España. Hemos roto el candado del 78. con y por vos...	@davidc_01 No, habéis dado carta blanca a Rajoy para daros mandanga. No sois independientes, estéis intervidos	1
RT @Rajocavi: Vuelvo y digo: Victor Cantillo es un manual de cómo jugar simple y efectivo. Lo mejor es que lo hace con clase, con categoría...	@Rajocavi Es como cuando estás viendo el partido en tv y dices allá allá está solo y justo el man la pone donde quieres es un crack!!	1
RT @planwac: Si se acalla la crítica en una Revolución, se muere la Revolución. El camino es el debate de ideas, la construcción colectiva...	@planwac Hace rato se calló la crítica, extraño darse cuenta ahora. Otros tiempos estuvieramos viviendo si el poder oficial hubiese escuchado	1
RT @pburelli: Los arrestados por ahora son: Nervis Villalobos Cárdenas, Luis Carlos de León, César Rincón Godoy y Rafael Reiter. Los próxi...	@pburelli @todoterenodx4 Así es	1
RT @Carpe6763Real: Se ha proclamado la independencia de Catalunya de España. Es más legal guerra a los españoles y llevarse la pasta a...	@Carpe6763Real @bea211m Cuando haces esa comparación, echas por tierra a la independencia	1
RT @pardodevera: Hay una manifestación violenta de ultras delante de Catalunya Radio, informan los compañeros/as, algunos encerrados/as den...	@pardodevera Es la mayoría silenciosa... los de siempre	1
RT @pardodevera: Hay una manifestación violenta de ultras delante de Catalunya Radio, informan los compañeros/as, algunos encerrados/as den...	@pardodevera Siempre los mismos...	1
RT @policia: Ayer, hoy y siempre defenderemos la ley, la convivencia y los derechos de todos los españoles. Unidos frente a la adversidad...	@policia VIVA ESPAÑA!!! Eses y su @policia y @guardiacivil orgullosos estoy de ellos	1
RT @rodrigoavea: Que genial que @sebastianpinera vaya a arreglar el Transantiago. Ojalá hubiera sido Presidente antes para tenerlo solucio...	@rodrigoavea @oposicionpinera @sebastianpinera Lo va arreglar si es un buen negociador para él o para alguno de los suyos pero ni cagando va a pensar en el tu bienestar. Tengalo presente	1
RT @elespanolcom: Societat Civil: "Hay que responder en la calle el domingo a este ataque a la democracia" https://t.co/tqz1Jcs4Wi?amp=1	@elespanolcom @Albert_Rivera @AlbertoLardies Jajajajaja sois una banda	1
RT @pablocasado: El día de la infamia no quedará impune. Los delincuentes sediciosos acabarán en la cárcel. Cataluña no es independiente v...	@pablocasado Tu y tu puta banda si que sois unos delincuentes y corruptos además de vividores de mierda !! dejar España ya	1
RT @pablocasado: El día de la infamia no quedará impune. Los delincuentes sediciosos acabarán en la cárcel. Cataluña no es independiente v...	@pablocasado ,será en tu pensamiento. Hace tiempo que Catalunya no es España. Q es que nos vas a fusilar a todos?	1

Table A41: Examples of labeled data: Twitter (6)

RT @pablocasado_: El día de la infamia no quedará impune. Los delincuentes s	@pablocasado_ Si señor🙄🙄🙄🙄🙄🙄🙄🙄🙄eseaseses	1
ediciosos acabarán en la cárcel. Cataluña no es independiente y...		
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, much	@Camila_Cabello @LatinAMAs Te amo	1
as gracias por tenerme en el show lo disfruté mucho #latinamas20...		
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, much	@Camila_Cabello @LatinAMAs Estoy tan orgullosa de ti bebe	1
as gracias por tenerme en el show lo disfruté mucho #latinamas20...		
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, much	@Camila_Cabello @LatinAMAs Yo y @wachademila queremos havana completa	1
as gracias por tenerme en el show lo disfruté mucho #latinamas20...	en español gracias	
RT @RaholaOfficial: VA PER TU, PARE!La República Catalana https://t.co/mbX9a	@RaholaOfficial Por Dios, qué razón, paren	1
RMD6D		
RT @IndicitaSelet: Yo tuve un presidente qué reconstruyó, paso a paso, el Esta	@IndicitaSelet Fue un gobierno excelente. Néstor siempre con el pueblo !!!!!	1
do. Yo tuve la suerte de vivir en la década ganada. https://...		
RT @IndicitaSelet: Yo tuve un presidente qué reconstruyó, paso a paso, el Esta	@IndicitaSelet Cómo te extraño!	1
do. Yo tuve la suerte de vivir en la década ganada. https://...		
RT @Sr.Dios: Puigdemont e Ignacio González organizando pachangas de solter	@Sr.Dios Mañana, me independizó del banco y dejo de pagar la hipoteca... po	1
os contra casados en Soto Del Real #DuiARV	que lo digo yo.	
RT @policia: Ayer, hoy y siempre defenderemos la ley, la convivencia y los dere	@policia Por favor, no estéis por mi, no hagáis más daño influenciados por los p	1
chos de todos los españoles. Unidos frente a la adversidad.	olíticos.	
@resistenciata Claro que sí!!!!!! Que oigan a esa gente que no está de acuerd	@resistenciata Ole ahí !!!!!!! @resistenciata venga vamos que nos vamos! M	1
o! 🙄🙄🙄🙄🙄🙄	ás fuertes que nunca !!!!!	
@resistenciata Claro que sí!!!!!! Que oigan a esa gente que no está de acuerd	@resistenciata Con letra o sin letra🙄🙄🙄🙄	1
o! 🙄🙄🙄🙄🙄🙄		
@resistenciata Claro que sí!!!!!! Que oigan a esa gente que no está de acuerd	@resistenciata @tonysandstowers todo preparado	1
o! 🙄🙄🙄🙄🙄🙄		
RT @alvaroojeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGO	@alvaroojeda80 Deapues de ver ésto, normal que se quieran independizar.	1
LES !!! VIVA ESPAÑA !!! https://t.co/qU52xnwdfJ		
RT @ensanro: La Declaración Unilateral Independencia es un error político que	@ensanro Vosotros sois los que no paráis de hacer crecer a la derecha español	1
hará retroceder a #Catalunya y fortalecerá a la derecha españ...	a.	
RT @ensanro: La Declaración Unilateral Independencia es un error político que	@ensanro @Caninator No nos quedan tantos no?	1
hará retroceder a #Catalunya y fortalecerá a la derecha españ...		
RT @ensanro: La Declaración Unilateral Independencia es un error político que	@ensanro España lo sabe mejor que nadie: es muy difícil vencer a movimiento	1
hará retroceder a #Catalunya y fortalecerá a la derecha españ...	independentista y en todo independencia hay sangre. #Catalunya	
RT @ElJeringasLoko: Así mi situación. https://t.co/hWwo0KxiCN	@ElJeringasLoko Así tu. @jennygarzaa	1
RT @Elaguujon_: A ver lo que hace mañana este gilipollas sin DNI. https://t.co/s	@Elaguujon_ Que prepare casi 11 euros el imbécil para sacarse uno nuevo. #ind	1
t1LNM7nZ6	ependentistaimbecil #tontodelculo #Articulo155	
RT @HogwartsTM: Los quiero Los tres. https://t.co/CPgsVOU04K	@HogwartsTM. Donde los venden ?	1
RT @albertmercade: #Alertaultra! Manifestació violenta a @CatalunyaRadio. htt	@albertmercade @VMMatterfilm @CatalunyaRadio Como siempre la violencia vie	1
ps://t.co/afSw2Ae7sz	ne de la derecha española	
RT @albertmercade: #Alertaultra! Manifestació violenta a @CatalunyaRadio. htt	@albertmercade @CatalunyaRadio Siempre igual, siempre los mismos...siempre	1
ps://t.co/afSw2Ae7sz	con la misma violencia porque no son capaces de hacer las cosas. EN PAZ	
RT @albertmercade: #Alertaultra! Manifestació violenta a @CatalunyaRadio. htt	@albertmercade @CatalunyaRadio Y como es que CatalunyaRadio no dice nad	1
ps://t.co/afSw2Ae7sz	a?	
RT @tonialba: A medida que vayan llegando te las enviamos. De momento: Bélg	@tonialba @tomijofresa De momento ningún país de esos apoya ni reconoce a	1
ica, Luxemburgo, Holanda, Finlandia, Estonia, Letonia... https://...	la catalunya independiente.No sigais engañando a la gente	
@tonialba Preneu-vos ja la pildora vermella.	@tonialba Claro que si	1
@tonialba Preneu-vos ja la pildora vermella.	@tonialba Si te crees esos apoyos eres muy tonto	1
RT @jon_mcenroe: Si no le comes las pelotas al independentismo eres un fascis	@jon_mcenroe @ElMolecula23 Jajajajaja, hombre, entre comer las pelotas y ten	1
ta	er claros niveles de poder y legitimida... https://t.co/K8rbnCyiNF	
RT @jon_mcenroe: Si no le comes las pelotas al independentismo eres un fascis	@jon_mcenroe Y viceversa	1
ta		
RT @MasQuePalabras: La verdad duele, la mentira mata... pero la duda, tortura.	@MasQuePalabras Me gusta	1
RT @MasQuePalabras: La verdad duele, la mentira mata... pero la duda, tortura.	@MasQuePalabras @Natalia_Malaga Y la amargura perdura.....aunque se apar	1
	ente lindura todo puede acabar en locura.	
RT @JosPastr: Los de Junts pel Sí y la CUP no pueden presentarse a las eleccio	@JosPastr Diría que la Cup hasta puede que saque algún diputado más si se pr	1
nes del 21 si quieren ser coherentes. Pero eso es suponer dem...	esenta	
RT @JosPastr: Los de Junts pel Sí y la CUP no pueden presentarse a las eleccio	@JosPastr @PeetarOtul Es que no sé cómo no están ilegalizados ya	1
nes del 21 si quieren ser coherentes. Pero eso es suponer dem...		
RT @JosPastr: Los de Junts pel Sí y la CUP no pueden presentarse a las eleccio	@JosPastr Esperemos que parte de ellos se nieguen a votar... Estaremos encant	1
nes del 21 si quieren ser coherentes. Pero eso es suponer dem...	ados.	
RT @JosPastr: Los de Junts pel Sí y la CUP no pueden presentarse a las eleccio	@JosPastr Cuando han sido coherentes?	1
nes del 21 si quieren ser coherentes. Pero eso es suponer dem...		
RT @JosPastr: Los de Junts pel Sí y la CUP no pueden presentarse a las eleccio	@JosPastr Ojala. Hay que extirpar este tumor, bueno no es ni tumor, que eso e	1
nes del 21 si quieren ser coherentes. Pero eso es suponer dem...	s serio, es como un grano en el culo que hay que quitar. 🙄🙄	
RT @JosPastr: Los de Junts pel Sí y la CUP no pueden presentarse a las eleccio	@JosPastr ya ha dicho que no se presentan	1
nes del 21 si quieren ser coherentes. Pero eso es suponer dem...		
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el golpe, s	@Albert_Rivera Os quiero os amo os adoro sois los mejores gracias por existir	1
e aplica el 155 CE y el Estado garantizará elecciones democráti...		
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el golpe, s	@Albert_Rivera Que elecciones limpias con un gobierno corrupto y fascista. Las	1
e aplica el 155 CE y el Estado garantizará elecciones democráti...	umas os undiran a la miseria. A ti Rivera y a los tuyos.	
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el golpe, s	@Albert_Rivera Claro que si...Ay si te dejaran...	1
e aplica el 155 CE y el Estado garantizará elecciones democráti...		
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el golpe, s	@Albert_Rivera Parace mentira que no sepas que lo que hace falta en Cataluña	1
e aplica el 155 CE y el Estado garantizará elecciones democráti...	ed un 155 largo y duro y no elecciones	
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprendido a	@julia_otero La pena es que la pregunta no puede ser independencia SI o NO.	1
todos. Esto no figuraba en las previsiones de analista algun...	La gente q siente la política como algo práctico votaría y saldría NO reparador	

Table A42: Examples of labeled data: Twitter (7)

RT @julia_oter: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprendido a todos. Esto no figuraba en las previsiones de analista algun...	@julia_oter ¡¡¡se te ha roto España !!! Qué malos sois...	1
RT @julia_oter: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprendido a todos. Esto no figuraba en las previsiones de analista algun...	@julia_oter Vaya... esta vez quizás sean los independentistas quienes no vayan a votar...	1
RT @julia_oter: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprendido a todos. Esto no figuraba en las previsiones de analista algun...	@julia_oter Como tú dijiste en una entrevista, esas elecciones tenían que haber tenido lugar antes; tenías razón. Pero Rajoy ha estado IMPECABLE	1
RT @julia_oter: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprendido a todos. Esto no figuraba en las previsiones de analista algun...	@julia_oter Grave error; esto no es algo que competa solo a los catalanes	1
RT @julia_oter: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprendido a todos. Esto no figuraba en las previsiones de analista algun...	@julia_oter A Rajoy le va las elecciones en Navidad	1
RT @julia_oter: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprendido a todos. Esto no figuraba en las previsiones de analista algun...	@julia_oter Esta vez el gobierno ha movido ficha y ha sido agil. Unas elecciones deberían ser una buena noticia para todos los demócratas	1
RT @CancinoAbog: Cual injuria ? Decirle asesino? Delincuente de lesa humanidad? Usted lo es! https://t.co/uVlxsyCkm	@CancinoAbog @jsatrich_farc si usted no es asesino ni terrorista, ni secuestrador, que somos los demás que solo vivimos dentro de la legalidad?	1
RT @edufeio: El parlamento se limpia un poco https://t.co/vZ2sfD8QBd	@edufeio Que buena noticia los montos afuera	1
RT @PabloMM: Pues al final los catalanes han hecho caso a muchos españoles y se han ido a su "puto país".	@PabloMM @eowynonfire Nos hemos quedado!, solo se irán los que no quieren estar. Nosotros nos quedamos. Cambia mucho!	1
RT @WaistedD: ¿Qué pasa si ves Stranger Things sin publicarlo en alguna red social?...¿te matan o que?	@WaistedD @chicabinaria Te pueden pasar cosas extrañas	1
@FonsiLoaiza @CatalunyaRadio Como se puede ir de digno periodista y decir que esa ha sido la única violencia? Eres un cobarde	@FonsiLoaiza @CatalunyaRadio Se responsable y no provoques ...	1
@FonsiLoaiza @CatalunyaRadio Como se puede ir de digno periodista y decir que esa ha sido la única violencia? Eres un cobarde	@FonsiLoaiza @CatalunyaRadio La violencia tiene muchas formas.	1
RT @TNFamosos: Mica Viciconte se animó a un primerísimo primer plano sin maquillaje https://t.co/YJAK2Kpe3c	@TNFamosos Es hermosa no necesita pintarse como una puerta!	1
RT @TNFamosos: Mica Viciconte se animó a un primerísimo primer plano sin maquillaje https://t.co/YJAK2Kpe3c	@TNFamosos @sabrincanooo Es hermosa sin/con maquillaje 💕	1
RT @TNFamosos: Mica Viciconte se animó a un primerísimo primer plano sin maquillaje https://t.co/YJAK2Kpe3c	@TNFamosos Es hermosa 😊	1
RT @TNFamosos: Mica Viciconte se animó a un primerísimo primer plano sin maquillaje https://t.co/YJAK2Kpe3c	@TNFamosos @MicaViciconte Belleza Natural la de ella, con maquillaje y sin. Hermosa	1
RT @AntonioMaestre: Los fascistas hoy saldrán de caza en Cataluña. Que se cuido la buena gente porque serán muy complicados.	@AntonioMaestre Especifica que fascistas, que no lo tengo muy claro quienes son.....	1
RT @kurioso: Hacer una votación secreta en el parlament en la que solo participas tú es como ponerte guapo para hacerte una paja	@kurioso @velascojmb Qué bueno!	1
RT @kurioso: Hacer una votación secreta en el parlament en la que solo participas tú es como ponerte guapo para hacerte una paja	@kurioso Muy bueno jaja..	1
RT @DJMaRiiO_90: Luego si me paso por casa de @Pumuscor podría hacer directo, luego os decimos 📺	@DJMaRiiO_90 @Pumuscor Suerte	1
RT @luipardo: Feijóo manda un mensaje unionista a Cataluña. Lo que pasó después te sorprenderá... https://t.co/ixdGJZV8RV	@luipardo @CarodRoviraLL Pues menos mal que alardean de respetar las lenguas y cultura de las diferentes comunidades autónomas !!	1
RT @luipardo: Feijóo manda un mensaje unionista a Cataluña. Lo que pasó después te sorprenderá... https://t.co/ixdGJZV8RV	@luipardo @Clio_deries jajajajaja.. tienen lo que se merecen, años sembrando odio, que esperan ahora?	1
RT @luipardo: Feijóo manda un mensaje unionista a Cataluña. Lo que pasó después te sorprenderá... https://t.co/ixdGJZV8RV	@luipardo Que raro!!!!	1
RT @thayspenalver: Ultima carta abierta a la MUD, por Thays Peñalver https://t.co/Gc1MbSwjrs	@thayspenalver @PerezGuadarrama Muy interesante, sin esgruirse a sabido to, dispuesta a confrontar con los 'políticos' la verdades (?) periodísticas, no creer ciegamente en gurúes que nos han traídos estos males. Es fácil enlodar, pero partirse el... es otra cosa. Enhorabuena! Saludos a los tuyos!	1
RT @Pablo_Iglesias_: La DUI no es legitima, pero la cuestión es: ¿por qué un 48% de catalanes quiere irse de España? La respuesta está en e...	@Pablo_Iglesias_ La pregunta que nunca tendrá respuesta es porque dejas te que Rajoy gobernara #politicolowcost	1
RT @Pablo_Iglesias_: La DUI no es legitima, pero la cuestión es: ¿por qué un 48% de catalanes quiere irse de España? La respuesta está en e...	@Pablo_Iglesias_ Si, la culpa es del chachachá. Pero gente de la mierda que tú representas votó a favor.	1
RT @IgnatiusFarray: Las élites de la burguesía catalana ya tienen lo que quieren: su propio Estado para montar su empresa, igual que el PP...	@IgnatiusFarray Pregunta, esto es comedia o opinión?	1
RT @IgnatiusFarray: Las élites de la burguesía catalana ya tienen lo que quieren: su propio Estado para montar su empresa, igual que el PP...	@IgnatiusFarray ¿Las élites que se piran a España?	1
RT @Cumparini: Me topé a Kast saliendo de Santiago por el finde largo https://t.co/kNL7depNON	@Cumparini Jajajajajaja buenísima 😊	1
RT @Cumparini: Me topé a Kast saliendo de Santiago por el finde largo https://t.co/kNL7depNON	@Cumparini No lo puedo creer, que horror!!	1
RT @maceirag: Galiza con Cataluña #Solidariedade Galiza amb Catalunya #Solidaritat #HolaRepublicaCatalana @llenguacat @omnium @assemblea...	@maceirag @omnium @llenguacat @assemblea @elciemen Que feliz me hace ver esto!	1
RT @maceirag: Galiza con Cataluña #Solidariedade Galiza amb Catalunya #Solidaritat #HolaRepublicaCatalana @llenguacat @omnium @assemblea...	@maceirag @omnium @llenguacat @assemblea @elciemen Mil gracias!	1
RT @maceirag: Galiza con Cataluña #Solidariedade Galiza amb Catalunya #Solidaritat #HolaRepublicaCatalana @llenguacat @omnium @assemblea...	@maceirag @llenguacat @omnium @assemblea @elciemen Muchas gracias	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @IndEnCastellano @CatalunyaRadio Gracias por explicarlo. L o hacemos nosotros y no se lo creen.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @IndEnCastellano @CatalunyaRadio Son así...	1
RT @olitorres10: Momento inolvidable. 📺 https://t.co/xbJ2Xqn7K	@olitorres10 El gol más rápido del Vicente CalderónOLIVER CRACK. Historia del @Atleti	1
RT @PPPopular: ► @marianorajoy explica las primeras medidas que se aplican del 155 📺 https://t.co/XPmmvqdLn4	@PPPopular @PEPEROES1972 @marianorajoy Os equivocáis.	1
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices. La seguridad de los justos. el sufrimiento de los humildes. #Né...	@marianorecalde Sobre todo el sufrimiento! De los humildes	1
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices. La seguridad de los justos. el sufrimiento de los humildes. #Né...	@marianorecalde Dejade de embromar, ni vos te crees eso.	1
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices. La seguridad de los justos. el sufrimiento de los humildes. #Né...	@marianorecalde Ok. Lo voy a recordar, champion. Y vos recorda que #No VuelvenMas	1

Table A43: Examples of labeled data: Twitter (8)

RT @marianorecalde: Te recordamos siempre junto a la risa de los felices.la seguridad de los justos.el sufrimiento de los humildes. #Né...	@marianorecalde Que lindo saber qué es lo que los está esperando.#NestorVive e n cada calabozo en el que está preso uno de sus secuestrados.	1
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices.la seguridad de los justos.el sufrimiento de los humildes. #Né...	@marianorecalde Ehhhhhhh????? @ @ @ @ @ Buenísimo!!! 👍👍👍👍👍👍👍👍👍👍	1
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices.la seguridad de los justos.el sufrimiento de los humildes. #Né...	@marianorecalde VERGUENZA ESCUCHAR A PERSONAS COMO USTED !!! Solo eso porque no quiero ser MALEDUCADA !!! 😡😡😡😡😡	1
RT @1Eme_: Sitios donde se reconoce la República Independiente de Cataluña#Du iARV #IndependenciaCatalunya https://t.co/P7laZND5LA	@1Eme_ ¿Hablas por todo el mundo? Que huevos tienes.	1
RT @mmendizabal1: Votar en secreto no es de valientes	@mmendizabal1 @13mininos Y no es legal!!	1
RT @mmendizabal1: Votar en secreto no es de valientes	@mmendizabal1 Manipular en las ondas, tampoco!	1
RT @mmendizabal1: Votar en secreto no es de valientes	@mmendizabal1 Pegar a viejas tampoco	1
RT @mmendizabal1: Votar en secreto no es de valientes	@mmendizabal1 Para nada, que de la cara el pueblo. Políticos...	1
RT @mmendizabal1: Votar en secreto no es de valientes	@mmendizabal1 Es de cobardes!!!	1
RT @AndresFelipe: ¿O sea que Santrich, culpable de asesinato de mi madre, me pue de denunciar por decirle asesino? Como va el país el tribun...	@AndresFelipe Qué tal este izquierdista	1
RT @AndresFelipe: ¿O sea que Santrich, culpable de asesinato de mi madre, me pue de denunciar por decirle asesino? Como va el país el tribun...	@AndresFelipe Es que en este país los miserables asesinos están gobernando	1
RT @AndresFelipe: ¿O sea que Santrich, culpable de asesinato de mi madre, me pue de denunciar por decirle asesino? Como va el país el tribun...	@AndresFelipe @MariaFdaCabal Este criminal creo que está desfasado, no soy abo gado, creo que aquí no aplica la tutela.	1
RT @IzcaNacional: Viva la República catalana!La hora de los pueblos#RégimenODE mocraciaGracias!	@IzcaNacional @assemblea @omnium @sumate_asoc @cupnacional @JuntsPelSi Muchas gracias	1
RT @IzcaNacional: Viva la República catalana!La hora de los pueblos#RégimenODE mocraciaGracias!	@IzcaNacional @assemblea @omnium @sumate_asoc @cupnacional @JuntsPelSi Gracias y no estáis solos!!	1
RT @IzcaNacional: Viva la República catalana!La hora de los pueblos#RégimenODE mocraciaGracias!	@IzcaNacional @assemblea @omnium @sumate_asoc @cupnacional @JuntsPelSi Gracias de corazón! No esperaba menos de la tierra de mi Madre y Abuelos mater nos.	1
RT @IzcaNacional: Viva la República catalana!La hora de los pueblos#RégimenODE mocraciaGracias!	@IzcaNacional @assemblea @omnium @sumate_asoc @cupnacional @JuntsPelSi Parece mentira que un grupo que se autodenomina izquierda defienda los naciona lismos, que no es mas que dar privilegios a una región quitando derechos a otras.	1
RT @IzcaNacional: Viva la República catalana!La hora de los pueblos#RégimenODE mocraciaGracias!	@IzcaNacional @assemblea @omnium @sumate_asoc @cupnacional @JuntsPelSi Gracias camaradas!!! VIVA ESPAÑA REPUBLICANA!!!! Viva su gente maravillosa!!!!	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Y siempre es así. Pero la mala publicitada es para los independentista.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @IndEnCastellano @CatalunyaRadio Gracias por explicarlo.Lo hacen os nosotros y no se lo creen.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Gracias por decir la verdad!! 🙏	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @IndEnCastellano @CatalunyaRadio Son así...	1
@juan_amorin: Imputaron a cuatro prefectos por la filtración de las imágenes de l a detención a Boudou. Y ojo que puede llegar a los func...	@juan_amorin Perdon pero no creo q llegue mas arriba	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio No creo en banderas queria decir ésto escribe lo que quiere.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Esos son los infiltrados de ultra derecha, no tienen conciencia ni respeto, solo asaltan y cobran su paga después de violentar y golpea r.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @Republic_Time @CatalunyaRadio No deciais que eran la mayoría si lenciosa? Pues le habéis tocado los huevos y bien, ahora van hacer ruido y con raz ón.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Por cierto no.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Gracias Fonsi. Hay mucha presión mediática para deshumanizar a todo un colectivo y así justificar su agresión. Te admiro por hacerl e frente	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí e s de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Lo de los coches de la Guardia Civil destrozados y el secuestro de la comitiva judicial fueron sueños, no te jode. ASCO DAS TÚ	1
RT @luis_sisepuede: Enhorabuena,ya no tendréis que elegir entre la papeleta azul, a naranja y la roja para votar a la derecha #L6ARVCat ht...	@luis_sisepuede @vmm7773 A partir de ahora, siempre juntos... Me da mucho mie do.	1
RT @luis_sisepuede: Enhorabuena,ya no tendréis que elegir entre la papeleta azul, a naranja y la roja para votar a la derecha #L6ARVCat ht...	@luis_sisepuede A ver si así, entre los tres...	1
RT @luis_sisepuede: Enhorabuena,ya no tendréis que elegir entre la papeleta azul, a naranja y la roja para votar a la derecha #L6ARVCat ht...	@luis_sisepuede @vmm7773 No se llama independentismo, se llama Republica Cat alana	1
RT @RominaGioPeron: Xavi intenta desescalar la tensión posando con un alcalde d e Ciudadans. https://t.co/HB8Guu0B3V	@RominaGioPeron Con uno? Con todos!!!	1
RT @Diego_Arria: A MUD(lodo)se le aplica Génesis 3:19 "Polvo eres, y en polvo te convertirás"	@Diego_Arria Así es, vamos a serlo ¿cuando se presentará la nueva alternativas ?	1
RT @Diego_Arria: A MUD(lodo)se le aplica Génesis 3:19 "Polvo eres, y en polvo te convertirás"	@Diego_Arria A fuego lento en el infierno.	1
RT @BelenBonnard: [👍👍] Una amiga es profesora en una escuela y hoy entró a la capilla y encontró esto en el altar: https://t.co/wCdaLduuFY	@BelenBonnard Ufff....se me apretó el corazón 🙏	1
RT @ladivaza: rt o tendrás 1000 años de mala suerte https://t.co/S7xyOvLOX	@ladivaza Como se te rompió eso? Se te cayó del avión o que vaina es?	1
RT @ladivaza: rt o tendrás 1000 años de mala suerte https://t.co/S7xyOvLOX	@ladivaza Todos los teléfonos de pedro tienen una maldición	1
RT @MicaSuarez12: NUEVO VIDEO, NUEVA DINÁMICA CON LA COMUNIDAD ❤️	Ft @BNToficial espero que les guste	1
RT @villalobossebas: te amo 🙏	@villalobossebas yo más más🙏🙏🙏🙏🙏🙏	1
RT @villalobossebas: te amo 🙏	@villalobossebas Yo te amo y te admiro🙏🙏🙏🙏	1

Table A44: Examples of labeled data: Twitter (9)

RT @villalobosbeas: te amo 🍷	@villalobosbeas TE AMO MÁS 🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Nosotros también !	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Te amo mucho demaciado. mas! 🍷🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Te amo muchísimo mas 🍷	1
@nelsonbocaranda: Rafael Ramirez montó en los pasillos frente al Consejo Seguridad ONU exposición acusando a la oposición de todos los m...	@nelsonbocaranda @Vshenko350 Hipócritas mentirosos! Gracias a Fios exist e material que demuestra todo lo contrario	1
@perezreverte: Sobre equiparación salarial PN y GC con otros cuerpos, como en tantas otras cosas, Mariano Rajoy ya prometía (y mentía) e...	@perezreverte Los @policia y @guardiacivil estamos cansados de que nos engañen y menosprecie.es hora de...	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Pensé que era DM. Yo también te amo.🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Te amo mas 🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas También te Amo	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Yo más porque en cada video me enseñas a nunca rendirme y seguir mis sueños. TE AMO nunca cambien porque eres una estupenda persona 🍷🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Yo Igual te amo baby @villalobosbeas	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Te amo más 🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Yo igual 🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Yo más. 🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Te amo mucho más 🍷	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Que lindo despertar con un te amo tuyo	1
@andahazi: Que D'Elia felicite a Sietecase y Tenembaum para mí es casi como ganar otro Martín Fierro.	@andahazi @2b1d811d257f453 Y por el contrario, si te felicita tienes que reflexionar a ver qué hiciste mal.	1
RT @villalobosbeas: te amo 🍷	@villalobosbeas Y yo a ti 🍷	1
@primavikinga: Hoy más que nunca me repito a mí misma que de todo se aprende... no hay errores en la vida... equivocarnos es vivir... ha...	@primavikinga Seamos felices todos	1
RT @babyqueenm: ¿Qué clase de A LITTLE TOO MUCH es este? Mis oídos han sido bendecidos #ARIASSHAWNMEÑDES https://t.co/1jNniEDnzN	@babyqueenm @GirlOnline5M Nunca la había escuchado, pero gracias a ti eso cambio	1
RT @EAristeguieta: Quienes han venido jugando a mantener a Maduro hasta el 2019, mientras aparentaban ser d oposición, son mas culpables que...	@EAristeguieta Si señor y que les revisen las cuentas y propiedades dentro y fuera del país.	1
RT @KatyTheChic: La gente no cambia, sólo se porta bien cuando quiere algo.	@KatyTheChic En el mejor de los casos podemos decir que uno puede mejorar como persona para sus seres queridos.	1
RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @angeleslagos Cierito, con su problema algunos politicos se escuchaban	1
RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @DeniseDresserG Claro!! Ya lo amenazaron!!	1
RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @DeniseDresserG Pues ahora que ya no será fiscal, puede revelar lo que sabe como cualquier ciudadano.	1
RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @DeniseDresserG Es decir que reculo!!!	1
RT @mcg1465: A esta perrita sus dueños NO la quieren más! SIN PALABRAS 🍷 Esta castrada, la trasladan donde alguien le de un lugarcito. Est...	@mcg1465 Su mirada lo dice todo...	1
RT @mcg1465: A esta perrita sus dueños NO la quieren más! SIN PALABRAS 🍷 Esta castrada, la trasladan donde alguien le de un lugarcito. Est...	@mcg1465 @fatcheves No la quieren mas??? Que gente de mierda tienen que sacar una ley para estos soreses	1
RT @mcg1465: A esta perrita sus dueños NO la quieren más! SIN PALABRAS 🍷 Esta castrada, la trasladan donde alguien le de un lugarcito. Est...	@mcg1465 Se nota pobre angel...yo acabo de adoptar a una gatita y es increíble la sensación de felicidad que me da este animal. quien pueda adoptela!	1
RT @mcg1465: A esta perrita sus dueños NO la quieren más! SIN PALABRAS 🍷 Esta castrada, la trasladan donde alguien le de un lugarcito. Est...	@mcg1465 @NicolettaArgY...a sus dueños. quien los quiere.???	1
RT @CGurisattiNTN24: Santos le decía al Fiscal que si no se aprueban beneficios para cultivadores de coca, le tocaría encarcelar a 100 mil...	@CGurisattiNTN24 A esta grosería le llaman presidente	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio @emparmoliner cuando quiera el pueblo	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio Eso debe ser inconstitucional. Debes convencer a PP, PSOE y Cs para cambiarla y permitirlo. Y cualquier otro intento acabará en tribunales.	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio Apoyo incondicional a ese referéndum.	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio @moneny68 Y yo estoy seguro que la República Catalana será el primer país en reconocer el nuevo estado español en forma de república. 🍷	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio Cuando se vote eliminar las autonomías...	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio Aquí, no hay, lo que tiene que haber !!!!	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio Me temo que los acontecimientos de estos días alejan esa posibilidad para unos cuantos años	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudaradio @emparmoliner Ya tardais	1
RT @Diego_Arria: A MUD[lo] se le aplica Génesis 3:19 "Polvo eres, y en polvo te convertirás"	@Diego_Arria Dr Arria, la gente aún no entiende que la MUD ha sido cómplice de esto. Me encanto este tuit. Saludos!!	1

Table A45: Examples of labeled data: Twitter (10)

RT @Diego_Arria: A MUD(lodo)se le aplica Génesis 3:19 "Polvo eres, y en polvo te convertirás"	@Diego_Arria profeta del desastre	1
RT @KatyTheChic: La gente no cambia, sólo se porta bien cuando quiere algo o.	@KatyTheChic En el mejor de los casos podemos decir que uno puede mejorar como persona para sus seres queridos.	1
RT @mcg1465: A esta perrita sus dueños NO la quieren más!SIN PALABRAS	@mcg1465 Su mirada lo dice todo....	1
ⓈEsta castrada, la trasladan donde alguien le de un lugarcito.Est... @annapacheco_: Hoy empieza el juicio a La Manada y hace un año escribí esto sobre grupos de WhatsApp. Igual cuesta admitirlo, pero algu...	@annapacheco_ Completamente de acuerdo.	1
RT @carnecrudadradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrudadradio Me temo que los acontecimientos de estos días alejan esa posibilidad para unos cuantos años...	1
RT @Diego_Arria: A MUD(lodo)se le aplica Génesis 3:19 "Polvo eres, y en polvo te convertirás"	@Diego_Arria Mientras mueren niños por hambre y enfermedades seguimos divididos el enemigo es quien dividió robo y sigue robando la lucha es x ven ezuela	1
RT @Diego_Arria: A MUD(lodo)se le aplica Génesis 3:19 "Polvo eres, y en polvo te convertirás"	@Diego_Arria Dr Arria, la gente aún no entiende que la MUD ha sido cómplice de esto. Me encanta este tuit. Saludos.	1
RT @GustavoBolívar: Rodrigo Lara Jr amigo de corruptos y enemigo de la paz, sería hoy la vergüenza d su padre quien dio la vida por hacer d...	@GustavoBolívar @hayquejodersen Yo creo que @Rodrigo_Lara_ lo sabe ! Decepción de su padre.!	1
RT @sailorfrag: Imagínense que padre que en vez de enseñarles a las mujeres a "darse a respetar" les enseñarían a los hombres a respetar :-o	@sailorfrag Te recuerdo que tú vienes de la vulva de una mujer y ella te educó. Reclamale a ella.	1
RT @sailorfrag: Imagínense que padre que en vez de enseñarles a las mujeres a "darse a respetar" les enseñarían a los hombres a respetar :-o	@sailorfrag No pues ya vi porque eres puto desviado	1
RT @Elaguujon_: El repaso de Arimadas a los nZionalistas se ha escuchado hasta en Andorra. https://t.co/HGvYRhQlk	@Elaguujon_ Y en las empresas, banca, colegios e iglesias de Cataluña.	1
RT @VickyVillarruel: Cuando la CIDH dejó de representar a todos y discriminó a las víctimas del terrorismo - mi nota en @infobae - https://...	@VickyVillarruel @beltrandelrio @infobae Estos organismos "autónomos" sólo responden a una agenda muy clara que es el nuevo marxismo: menos humanismo y más totalitarismo	1
RT @rubencaravaca: ¿Nadie dice nada de la aprobación del CETA en el Senado? un golpe de estado real contra nuestra soberanía ¿Qué nos está...	@rubencaravaca A Cataluña ese voto ya no la vincula. ¿Qué vais a hacer vosotros?	1
RT @rubencaravaca: ¿Nadie dice nada de la aprobación del CETA en el Senado? un golpe de estado real contra nuestra soberanía ¿Qué nos está...	@rubencaravaca Pues que somos muy tontos, eso nos pasa	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok hay que lindo adoro a los bebés	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Que alegría!! va a ser uno de los bebés más queridos por los argentinos.❤	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Que alegría!! Te deseo mucho 💕	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok @soycrismorena Muchas felicidades!!!! Abrazo enorme	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicidades Marley!!	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicidades PAPÁ MARLEY!!! Que Dios los bendiga a ambos!!!	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicidades papá!!!	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicidades! Bienvenido Mirkol! 🥰🥰	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok FELICITACIONES DIOS OS BENDIGA MUCHISIMO... 🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok felicidades @cokemongo ya eres hermana mayor	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Vamos Marley felicitaciones!!! Se me caen las lágrimas por ponerme en tu lugar. Bendiciones!!!	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicitaciones!!!!	1
@marley_ok: Gracias!!!!	@marley_ok Que felicidad	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Me conmueve hasta las lágrimas. Muchas felicidades y tranquilo que vas a ser un buen padre!! 🙏	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Hermosos que vuelvan rápido los dos 🙏🙏🙏🙏	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicitaciones 🙏	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá. 51 cm. 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicitaciones... dios lo bendiga siempre🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏	1
RT @IgorMeltxor: ALERTA: nazis de caza por Barcelona aplicando su 155.#L6ARVCat #RepúblicaCatalana #IndependenciaCatalunya #NoPassaran #se...	@IgorMeltxor Nazis? Vosotros podeis sacar vuestra bandera y nosotros no? Quieras o no, esa bandera te sigue representando. Republica catalana dice ja ja	1
RT @IgorMeltxor: ALERTA: nazis de caza por Barcelona aplicando su 155.#L6ARVCat #RepúblicaCatalana #IndependenciaCatalunya #NoPassaran #se...	@IgorMeltxor Ojala les peguen bien fuerte	1
RT @ladivaza: rt o tendrás 1000 años de mala suerte https://t.co/57jxyOvLOX	@ladivaza @Camy_23Benitez Sigue estando mejor que el mio.	1
RT @joosearodriguez: 1.850.000 independentistas catalanes que han rajado su DNI al proclamar la independencia, a 10.60€ por renovar son 1...	@joosearodriguez @JorgeManzanoLag No eran mas de 2.000.000!!!!	1
RT @Santi_Vatt: El velorio turbina https://t.co/MZZm6QNxmN	@Santi_Vatt Te pones a pensar a dónde está subido el cajón y vos tipo...	1
RT @eclementen: El Estado opresor y franquista humilla a Cataluña poniendo urnas de verdad y obligando a los catalanes a votar con garantía...	@eclementen Tu eres periodista? Vergüenza para la profesión.Como se puede decir tantas tonterías? Dedicate al cante, que como periodista dejas mucho que desear	1
RT @eclementen: El Estado opresor y franquista humilla a Cataluña poniendo urnas de verdad y obligando a los catalanes a votar con garantía...	@eclementen @pabloharour Todos esos votos impresos que los nacionalistas querrán meter de 6 en 6 como la última vez Y urnas opacas llenas de votos que no volverán a ver	1
RT @jmdelalamo: Hay dos tipos de países: los que no reconocen la independencia de Cataluña y los que no saben lo que es Cataluña.	@jmdelalamo Calma. Hoy ya lo saben mas que ayer. Poco a poco.	1
RT @jmdelalamo: Hay dos tipos de países: los que no reconocen la independencia de Cataluña y los que no saben lo que es Cataluña.	@jmdelalamo mientras algunos lloraban de alegría debería haber saltado un o diciendo "inocente, inocente" jaja, es increíble el fanatismo de la gente	1

Table A46: Examples of labeled data: Twitter (11)

RT @ed000chan: Viernes de SPOILERS #NFAH3-3000 RT Pesadillas-4000 RT O SCURIDADA-5000 RT KILLER NIGHTVAMWAYS A DIVERTIRNOS!!!	RT @ed000chan A divertirnos	1
RT @ed000chan: Viernes de SPOILERS #NFAH3-3000 RT Pesadillas-4000 RT O SCURIDADA-5000 RT KILLER NIGHTVAMWAYS A DIVERTIRNOS!!!	RT @ed000chan Todo eso me suena a los nightmares a halloween xd <3	1
RT @FeedeeHerediaa: Fernet toda la vida no hay comparación el smirnoff es pa ponerse en pedo, el Fernet se disfruta con el alma y corazón...	RT @FeedeeHerediaa @EnzoLolla Con el alma y corazón. Amen	1
RT @protestona1: Yo creo que el día de la infamia fue aquel 11M que el PP realizó 12 operaciones de blanqueo mientras sacábamos a los muert...	RT @protestona1 @xmalban Pero que sabras tu lo que es la infamia? infamia fue ver como la policia fascista pagaba a mi padre y a otros obreros a las puertas d e Citroen.	1
RT @PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de #EnergiasRenovables	RT @PRODUCCION_ARG Ya era hora que pasen cosas buenas para. El país / el crecimiento nos da mas posibilidades para todos !!	1
RT @PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de #EnergiasRenovables	RT @PRODUCCION_ARG Brillante!!!	1
RT @PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de #EnergiasRenovables	RT @PRODUCCION_ARG Felicidades!! Haciendo lo que hay que hacer!! lo están haciendo!! GRACIAS!!	1
RT @PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de #EnergiasRenovables	RT @PRODUCCION_ARG Siempre avanzando	1
RT @PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de #EnergiasRenovables	RT @PRODUCCION_ARG @dr_kuato Esto es muy bueno para nosotros	1
RT @PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de #EnergiasRenovables	RT @PRODUCCION_ARG Excelente !!!!	1
RT @SusodeToro1: Fascistas están asaltando la radio catalana. Si los medios es pañoles no la están denunciando entonces es que son cómplices	RT @SusodeToro1 No esperes que Ferreras lo denuncie con la deriva que lleva. Hoy ha piragado a Manano R. Al final adelantará por la dcha. a Falangito	1
RT @saragarcia14: Hacen faltan personas con discursos como los de @nataliaferri Mujeres fuertes y libres siempre	RT @saragarcia14 @marta_granero @nataliaferri @marta_granero verguenza so tenia que dar a las de la tele, que sois las que enseñais a las niñas de 12 a hacer todo lo que la tipa este dice	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	RT @FonsiLoaiza @CatalunyaRadio Que hipocrita eres! La del bolazo en el ojo iba a pasar tranquilamente por las ramblas un mosso se tropezó y le dio un bolazo, no?	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	RT @FonsiLoaiza @CatalunyaRadio Y además son cuatro.	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	RT @FonsiLoaiza @CatalunyaRadio Tu odio te impide ver siempre la realidad.	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok felicidades @marley_ok es muy grandote el bebe, Dios los bendiga	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok Felicidades Marley! Lo lograste!	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok Que bueno, me alegro muchísimo!!!! Ahora a no dormir por las noches jajajaa	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok @soycismorena Felicidades Marley	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok Muchas felicidades Que la disfrutes	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok Que felicidad mas grande!!!! Lo mejores deseos y a disfrutar de éste gran amor!!!	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok Bienvenido	1
RT @Gluogo23: SE VIENE EL CLÁSICO SUE E @TigresOficial SE LO IMPO ANTE Q ES ESTE JUEGO PARA TODA LA AFICIÓN FELINA @LIBRESYLOKOS	RT @Gluogo23 @TigresOficial @LIBRESYLOKOS Excelentes recuerdos tuyos con TigresOficial	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok Felicidades Marley muchas bendiciones. besos	1
RT @marley_ok: Nació Mirkol!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació con 7 semanas, necesita un poco aún de cuidados, pero muy felices!...	RT @marley_ok Que emoción felicitaciones marley	1
RT @AgusPrados: Así estamos, desafiando y creyendo q somos superiores a una autoridad de formación profesional. No estudiemos más! Seamos u...	RT @AgusPrados Que se puede esperar https://t.co/sbGadnGcIV	1
RT @AgusPrados: Así estamos, desafiando y creyendo q somos superiores a una autoridad de formación profesional. No estudiemos más! Seamos u...	RT @AgusPrados @Ricardodepalos Tiene una foto de una nena de 5 años fumando. Es terrible troll o terrible cabeza de termo, como sea tiene retraso mental.	1
RT @jcreyes: Si vota 60% del padrón, Piñera baja del 30%, por eso están asustados x el llamado a votar q hace el gobierno.Ojalá lo entien...	RT @jcreyes @chileconvoa Eres adivino y crees que la mayoría de los jóvenes son p...? por qué no?	1
RT @renfrancis: Ya casi es Noviembre y yo signo sin ti.	RT @renfrancis Ya somos dos! Yo signo sin conocer a mi bello y a mis CNCO's	1
RT @renfrancis: Ya casi es Noviembre y yo signo sin ti.	RT @renfrancis Y casi es año nuevo y yo no he podido abrazar a los chicos	1
RT @renfrancis: Ya casi es Noviembre y yo signo sin ti.	RT @renfrancis Dile eso a los chicos de mi parte	1
RT @montanertwitter: Después d unos cuantos años en este negocio,concluyó, q el primer enemigo de nuestro talento, es nuestro ego...Nada, e...	RT @montanertwitter Es verdad, y entre mas grande es el ego, mayor es nuestro enemigo	1
RT @montanertwitter: Después d unos cuantos años en este negocio,concluyó, q el primer enemigo de nuestro talento, es nuestro ego...Nada, e...	RT @montanertwitter Sin dudas !!!	1
RT @montanertwitter: Después d unos cuantos años en este negocio,concluyó, q el primer enemigo de nuestro talento, es nuestro ego...Nada, e...	RT @montanertwitter motaner es un persona muy famosa y cantante muyconocido el colombia y le mundo	1
RT @elephantiny: Si no gritaste cuando apareció Micaela con este peinado NO SE QUE HACES EN ESTE FANDOM #HallowenYoutuber https://t.co/mX6...	RT @elephantiny @MicaSuarez12 Jaja si yo grite mal	1
RT @elephantiny: Si no gritaste cuando apareció Micaela con este peinado NO SE QUE HACES EN ESTE FANDOM #HallowenYoutuber https://t.co/mX6...	RT @elephantiny @MicaSuarez12 Exacto!!!!!!	1
RT @apguayaz: Nuestro Presidente Vitaleio del Movimiento 35PAIS, @MashiRafael, envia un mensaje a toda la militancia del @Guayaz, @marce...	RT @apguayaz @35PAIS @MashiRafael @marcelaguinaga Aquí habla tanto medio cre justiciero y seguro creen en Dios tal parece que es un Dios d maldad por que? Hablan cualquier cosa	1
RT @CUBONoticiasMX: ¿Y mira @alfredodelmazo así "tratan" a los reos del Pena I. Neza Bordo ¿Algo qué decir?	RT @CUBONoticiasMX @TapiaFernanda @alfredodelmazo Ahí deberían de estar lo s gobernadores rateros y darles la misma bienvenida.	1
RT @MasQuePalabras: "Humildad" El valor más hermoso del ser humano, pero también el más escaso.	RT @MasQuePalabras Entre más éxito consigas más humilde se tiene que ser. Es un valor muy importante que las personas lo están perdiendo	1
RT @Saul_Craviotto: Creo que ha llegado la hora de pronunciarme. No me podí a llamar más. #Catalunya #España https://t.co/BxCxCMwIV	RT @Saul_Craviotto @policia Estamos muy orgullosos de ti tanto por tus troncos d epaños como por tu arrojo personal un abrazo desde zaragoza	1

Table A47: Examples of labeled data: Twitter (12)

RT @Saul_Craviotto: Creo que ha llegado la hora de pronunciarme. No me po dia llamar más. #Catalunya #España https://t.co/BxCx8mVwFw	@Saul_Craviotto Estamos muchos millones de españoles con vosotros. Un millón de gracias...
RT @Barrayco: La Cataluña ignorada por el nacionalismo https://t.co/osjM9 OKWv3	@Barrayco Este video es de actualidad o es de hace días.
RT @lopezobrador_: Claudio X. González, quien destapó a Meade y siempre me ataca, es un salinista megacorrupción que participó en el fraude e...	@lopezobrador_ Ay! Siempre hablando de los otros lo que tu, tus hijos y los de tu partido hacen con un cinismo desmedido.
RT @lopezobrador_: Claudio X. González, quien destapó a Meade y siempre me ataca, es un salinista megacorrupción que participó en el fraude e...	@lopezobrador_ Es tiempo de recuperar la República. Nos toca a los de abajo reconstruir nuestra patria. Defendamos al país todos con AMLO. Sonrían gana mos.
RT @jorgeramosnews: O sea, en México no pasa nada. Hay una clase política casi intocable, donde se cuidan unos a otros, y urge cambio tipo...	@jorgeramosnews Que encabronamiento... usted disculpe.
RT @jorgeramosnews: O sea, en México no pasa nada. Hay una clase política casi intocable, donde se cuidan unos a otros, y urge cambio tipo...	@jorgeramosnews Lo dudo
RT @jorgeramosnews: O sea, en México no pasa nada. Hay una clase política casi intocable, donde se cuidan unos a otros, y urge cambio tipo...	@jorgeramosnews como que no pasa nada, terribles huracanes, desastrosos t erremotos y una robadera de las donaciones y seguro veremos tiendas de ca mpaña por ai
RT @jorgeramosnews: O sea, en México no pasa nada. Hay una clase política casi intocable, donde se cuidan unos a otros, y urge cambio tipo...	@jorgeramosnews espero eso nos mueva de lo contrario nos va ir peor!!
RT @jorgeramosnews: O sea, en México no pasa nada. Hay una clase política casi intocable, donde se cuidan unos a otros, y urge cambio tipo...	@jorgeramosnews No hacemos nada seguimos sin hacer nada contra nuest os políticos ☹️
RT @Rubius: Youtube va un poco lento hoy 🐢 No se procesa bien el video	@Rubius Qué raro youtube y su lentitud
RT @Rubius: Youtube va un poco lento hoy 🐢 No se procesa bien el video	@Rubius Yo ya lo vi y no paso nada
RT @carlosmatareal: Lamentable fallecimiento para todos los venezolanos. Fre ddy Galavis se nos fué esta madrugada. Abrazo y buen viaje quer...	@carlosmatareal Lo siento mucho. Mis mas condolencias para las familias. Lo recuerdo de la novela Juan la Virgen.
RT @carlosmatareal: Lamentable fallecimiento para todos los venezolanos. Fre ddy Galavis se nos fué esta madrugada. Abrazo y buen viaje quer...	@carlosmatareal Que pena que diosito lo tenga en su gloria 🙏🙏🙏🙏🙏
RT @GuillermoBotero: @BancoRepublica reduce tasa de interés en 25 pb a 5 %. Están muertos del susto con desempeño económico. ¡Ojalá bancos l...	@GuillermoBotero @JuanLozano_R @BancoRepublica Analice inteligentem e
RT @AlvaroUribeVel: Retén terrorista. Farc en el Cauca, guardaron el uniforme y sacaron las armas que se supone habían entregado. https://t...	@AlvaroUribeVel Que descarados!
RT @LulaTigre: Ayudemos a Jorge. El quiere y necesita laburar. Tiene CV. Este es su numero 15.5973.1020. Hace 5 meses se quedo sin trabajo...	@LulaTigre @botedeltitanic1 Fuerza! Dios no te abandonará!!
RT @Lucas_Castel: Nadie encuentra su camino, sin haberse perdido varias vec es. https://t.co/vwqghcvAp1	@Lucas_Castel no sé por que
RT @GuillermoBotero: Primero suspendieron fumigación aérea y ahora quiere n despenalizar siembra de 3.8 hectareas de cultivos ilícitos. Dond...	@GuillermoBotero Tenemos que salir a las calles a derrocar a este sicopata n acro guerrillero!
RT @GuillermoBotero: Primero suspendieron fumigación aérea y ahora quiere n despenalizar siembra de 3.8 hectareas de cultivos ilícitos. Dond...	@GuillermoBotero Cuando apoyaron y eligieron a Juhampa.
RT @GuillermoBotero: Primero suspendieron fumigación aérea y ahora quiere n despenalizar siembra de 3.8 hectareas de cultivos ilícitos. Dond...	@GuillermoBotero El tren de la legalización es el verdadero muñeco de la paz ¡busque la participación del 8000 y los vera alineados con proposito claro...
RT @EPN: México no reconocerá la declaración unilateral de independencia d e Cataluña. Hacemos votos por una solución política y pacífica.	@EPN @LocraTapia Preocuparte por tu pueblo es lo que debes hacer y no p or otros países
RT @EPN: México no reconocerá la declaración unilateral de independencia d e Cataluña. Hacemos votos por una solución política y pacífica.	@EPN Tu no, pero yo sí ♥️
RT @EPN: México no reconocerá la declaración unilateral de independencia d e Cataluña. Hacemos votos por una solución política y pacífica.	@EPN Pues si no te estan pidiendo opinion jajajaja
RT @guspavan: Mi vecina de abajo... https://t.co/SeVEiOIXa	@guspavan @pablitollescano @FedeCastronovo No es tu vecina?
RT @guspavan: Mi vecina de abajo... https://t.co/SeVEiOIXa	@guspavan @SinCodigoTV @macmacaco es tu vecina??? Jajajaja
RT @Wil_Scorpion: #MeAsustoCuandoMeDicen que la corrupción no existe. C uántos años de cárcel ven en este cartón de @monerorape 🙄 https://t...	@Wil_Scorpion @monerorape Es pura percepción...
RT @bysTaXx: Aquí con @ByRubii9 ☹️ https://t.co/GY01xLFas1	@bysTaXx @ByRubii9 Ahi tu todo feliz con el globo como un niño🤡🤡
RT @gerardotc: Esta noche en CAT hay nazis atacando, pero llamarlos nazis s eria un lio pq ya habíamos quedado en que los nazis eran los de...	@gerardotc Creo que lo dice en plan irónico, pero quizás me equivooco
RT @gerardotc: Esta noche en CAT hay nazis atacando, pero llamarlos nazis s eria un lio pq ya habíamos quedado en que los nazis eran los de...	@gerardotc Si, menudos cabrones. Les dicen que no son de donde eres toda la vida y se enfadan. Vaya nazis.
RT @gerardotc: Esta noche en CAT hay nazis atacando, pero llamarlos nazis s eria un lio pq ya habíamos quedado en que los nazis eran los de...	@gerardotc Más bien parece el principio de la desnazificación.
RT @gerardotc: Esta noche en CAT hay nazis atacando, pero llamarlos nazis s eria un lio pq ya habíamos quedado en que los nazis eran los de...	@gerardotc Los nazis no existen, son los padres.
RT @gerardotc: Esta noche en CAT hay nazis atacando, pero llamarlos nazis s eria un lio pq ya habíamos quedado en que los nazis eran los de...	@gerardotc Son fascistas/neonazis sin ninguna duda, yo lio ninguno, hay que pararles los pies todos.
RT @gerardotc: Esta noche en CAT hay nazis atacando, pero llamarlos nazis s eria un lio pq ya habíamos quedado en que los nazis eran los de...	@gerardotc Es que votar es muy nazi. Ser agredido también. ¿No lo sabiais? Yo tampoco.
RT @Pablo_Iglesias_: 3. Seguiremos defendiendo el diálogo y un referéndum l egal y pactado como mejor solución a la crisis en Catalunya	@Pablo_Iglesias_ Y dale con el referendun!!; que mania!! ¿ por que te intetes a tanto?
RT @Pablo_Iglesias_: 3. Seguiremos defendiendo el diálogo y un referéndum l egal y pactado como mejor solución a la crisis en Catalunya	@Pablo_Iglesias_ No hay diálogo con los golpistas, bien lo sabia tu querida re pública.
RT @Pablo_Iglesias_: 3. Seguiremos defendiendo el diálogo y un referéndum l egal y pactado como mejor solución a la crisis en Catalunya	@Pablo_Iglesias_ Eres un inútil. Mejor seria que te dedicaras a otra cosa.
RT @BelenBonnard: [👉👈] Una amiga es profesora en una escuela y hoy ent ró a la capilla y encontró esto en el altar: https://t.co/wCdaduFY	@BelenBonnard @juliansalcedo Amén Amén . Que así sea
RT @astros: LA PIÑA! 🍍🍍🍍 https://t.co/BBMHdOjHjr	@astros 🍍🍍🍍🍍🍍🍍🍍🍍 Viva La Piña!!!!
RT @jtazmuri: O los ASESORES que ayudaron a la Josefa a desfondar el Munic ipio de Providencia?	@jtazmuri @orangni si estos mediocres creen que pueden es populismo. Los mejores cerebros no han podido
RT @susana_zori: Hoy hace un año q falleció 🍂 mi hijo Pablo atropellado y a bandonado por un borracho al volante. Ahora quiero cárcel para él...	@susana_zori @UsinadJusticia Que terrible, te admiro la fuerza.

Table A48: Examples of labeled data: Twitter (13)

RT @PiensaPrensa: [LO ULTIMO] PIÑERA: "Si la Presidenta cierra Punta Peuco, voy a volver a abrirla". https://t.co/e7lau5Hbdn	@PiensaPrensa que asco de persona es ese Piñera	1
RT @PiensaPrensa: [LO ULTIMO] PIÑERA: "Si la Presidenta cierra Punta Peuco, voy a volver a abrirla". https://t.co/e7lau5Hbdn	@PiensaPrensa Es muy estúpido este delincuente!!	1
RT @ziberial: Casi me echan del bar de la risa https://t.co/R7t39uKN8K	@ziberial lo peor que no estaba mirando 😞	1
RT @ziberial: Casi me echan del bar de la risa https://t.co/R7t39uKN8K	@ziberial Igual al final no se que de qué se rie	1
RT @ziberial: Casi me echan del bar de la risa https://t.co/R7t39uKN8K	@ziberial No lo habia visto	1
RT @ziberial: Casi me echan del bar de la risa https://t.co/R7t39uKN8K	@ziberial Es genial	1
RT @christophervele: #CNCOWners 4/4 🍷❤️ son el mejor fandom del mundo! 🍷🍷🍷🍷	@christophervele Y tú el mejor idolo	1
RT @christophervele: #CNCOWners 4/4 🍷❤️ son el mejor fandom del mundo! 🍷🍷🍷🍷	@christophervele Hoy y Siempre , lo Somos❤️ mi bebé💕	1
RT @christophervele: #CNCOWners 4/4 🍷❤️ son el mejor fandom del mundo! 🍷🍷🍷🍷	@christophervele Felicitaciones @CNCOMusic por los premios	1
RT @christophervele: #CNCOWners 4/4 🍷❤️ son el mejor fandom del mundo! 🍷🍷🍷🍷	@christophervele Gracias les queremos y daríamos todo por ustedes	1
RT @marquito_vyro2: "como será que vive la gente que no le gusta el sushi" https://t.co/oHLS2CgTtI	@marquito_vyro2 Se me hace agua en la boca🍷😞	1
RT @chatoguedez: Según Falcón su derrota electoral y los +100 caído s son culpa de la oposición y no de la dictadura asesina y tramposa. Se...	@chatoguedez @Carlos_BolivarF @Carlos_BolivarF podrias pasar un cap ture de cómo se hace el trámite de la prórroga?? Por favor	1
RT @chatoguedez: Según Falcón su derrota electoral y los +100 caído s son culpa de la oposición y no de la dictadura asesina y tramposa. Se...	@chatoguedez y de quien es la culpa pues..? ustedes mandaron a eso s muchachos a la calle y no pararon la masacre..... https://t.co/VnPkIWUH9s	1
@dosogas actuo mejor yo q ustedes ahre	@dosogas No lo sé, parece falso	1
@JavierMMoure @jinsua49 Con esa vocesita de nenita, hacé shhhh y no jodas más. https://t.co/pRoROJKix3	@JavierMMoure Pobrecito, se enfrentó a la realidad....	1
@JavierMMoure @jinsua49 Con esa vocesita de nenita, hacé shhhh y no jodas más. https://t.co/pRoROJKix3	@JavierMMoure @ArielJob Que deje de joder . Basura humana	1
@lo_popular Pampita (a muerte)	@lo_popular @ReinaPampita Yo voto a pampita	1
@lo_popular Pampita (a muerte)	@lo_popular Amamos a Pampita.	1
RT @cncxmendes: Me muerdo de la vergüenza pero acá les dejo mi c over de #Mamita - @CNCOMusic #MamitaFanVideo https://t.co/KGHF4eN1aP	@cncxmendes @CNCOMusic Que bonito cantas y tocas la guitarra t e salio hermoso 😊😊😊	1
RT @yuyacst: Te amo.	@yuyacst Yo más💕💕	1
RT @yuyacst: Te amo.	@yuyacst Te amo mas❤	1
RT @yuyacst: Te amo.	@yuyacst Gracias ^^ ya lo sabia!	1
@isco_alarcon: Recuperando después del golpe del sábado, buenas se nsaciones y con ganas de derbi 2 2	@isco_alarcon Después es Piqué. Cada vez que hay Seleccion algún m adridista se borra. O quizás sea un mutuo acuerdo...	1
RT @SergioFVazquez: Otro clásico más y la Cato no esta en un gran momento deportivo pero siempre está el temple de cada jugador para poder...	@SergioFVazquez Que gran equipo! Estuve en el estadio y siempre m e acuerdo!! Como se quedaron esos chunchos espero mañana lo gane mos!!	1
RT @AlboranTeam: #Video Mágico @pabloalboran esta noche cantan do #Saturno en premios @LatinAMAs desde el @DolbyTheatre de LA	@AlboranTeam @pabloalboran @LatinAMAs @DolbyTheatre La ambie ntación del escenario perfecta!!👏👏 Brilló como nunca	1
RT @danilazarom: El director general de los Mossos, Pere Soler, se des pide por carta. Acata pues la destitución del Gobierno de la Nación....	@danilazarom Esto ya lo tenía más que escrito 😊😊	1
RT @danilazarom: El director general de los Mossos, Pere Soler, se des pide por carta. Acata pues la destitución del Gobierno de la Nación....	@danilazarom @Rosevelz Este tio Pere Soler ha sido un cancer en el c uerpo de los Mossos	1
RT @Patriciomonero: Selección infalible https://t.co/YT7MpMd51X	@Patriciomonero @ElChamuco Te felicito, ahora si ya puedo decir qu e eres un monero parejo	1
RT @mateodualdeok: toy morido https://t.co/PRsOUeWnos	@mateodualdeok Bien la chica por no cagarse de risa 🤔🤔🤔	1
RT @yFervazquez: Cri criminal 🤔	@yFervazquez Porque esa cara	1
@AlvaroUnibeVel: Petro creó 50 mil contratos burocráticos en Bogotá, quebró la ciudad y aspira a la Presidencia	@AlvaroUnibeVel Y usted???? Tanto que hizo.... se le olvidó ? y aún sig ue detrás del poder.... deje ya la avancia...	1
@malepichot: Gracias a la persona maravillosa que hizo esta genialida d	@malepichot Buenísimo y profundamente deprimente. Tanto por cam biar.	1
RT @yFervazquez: Cri criminal 🤔	@yFervazquez Tu estilo tu flow baby muy criminal	1
@XavierGoda: "A los presuntos violadores de Sanfermines, la acusaci ón les pide 22 años. A mi hijo, por una pelea de bar, le piden 50 añ...	@XavierGoda @emparmoliner Muchos ánimos y no parar de luchar!	1
@pccomponentes: SO EAZOI! Comenta con y sigue a PcComponentes y @OMENbyHP y muy importante rellena tus dat...	@pccomponentes @OMENbyHP Un fantastico equipo y un gran sorte o a todos. Buena suerte a t...	1
@Rainbook: Diganme que ustedes también se suben a un elevador y c antan: let's get together yeah yeah yeah	@Rainbook No, pero lo empezaré a hacer desde ahora.	1
RT @Checheperez: Jorge Glas defendiendo a Odebrecht en cadena na cional. Sabía hasta cuánto habían ofertado los otros pero ahora dice q ue él...	@Checheperez Gano beneficio para los bolsillos de el mismo y del tio.	1
RT @Checheperez: Jorge Glas defendiendo a Odebrecht en cadena na cional. Sabía hasta cuánto habían ofertado los otros pero ahora dice q ue él...	@Checheperez Entiendan entre amigos no se acusan, Odebrech está a poyando a la oposición. Porque será?	1
RT @Checheperez: Jorge Glas defendiendo a Odebrecht en cadena na cional. Sabía hasta cuánto habían ofertado los otros pero ahora dice q ue él...	@Checheperez Porque Odebrech acusa al vicepresidente y no dice nada de los contratos de anteriores gobiernos	1
RT @filmagicpy: ¿Querés ver al Dios del Trueno🌩 y llevarte un kit?;C on un RT ya participá! Tenés tiempo hasta el lunes 🤩#ThorRagnarok h...	@filmagicpy Quiero ganar el kit #ThorRagnarok mi Cl es 442	1

Table A49: Examples of labeled data: Twitter (14)

RT @LaurenJauregui: Lo que quieras si me dices Por Favor https://t.co/4eMm4egUw	@LaurenJauregui Besame Por Favor!	1
@Pontifex_es: Nada ni nadie puede oscurecer la luz que Cristo pone en el corazón y en el rostro de sus amigos.	@Pontifex_es Asi es! Bendiciones para ti Papa Francisco. Que Dios Te Regale Muchos Años de vida	1
RT @LaurenJauregui: Lo que quieras si me dices Por Favor https://t.co/4eMm4egUw	@LaurenJauregui Por favor sigueme 🙏	1
RT @LaurenJauregui: Lo que quieras si me dices Por Favor https://t.co/4eMm4egUw	@LaurenJauregui Por supuesto cuando quieras si me dices por favor	1
@erickfancy Vayan al video original y denle RT muchas gracias por publicar nuestro video!❤️	@erickfancy Lo mejor que vi la puta madre es genial jajaja	1
@erickfancy Vayan al video original y denle RT muchas gracias por publicar nuestro video!❤️	@erickfancy Gracias por compartir nuestro video 😊😊	1
@erickfancy Vayan al video original y denle RT muchas gracias por publicar nuestro video!❤️	@erickfancy vayan a darle a este ah https://t.co/JzDHRKKds1	1
@erickfancy Vayan al video original y denle RT muchas gracias por publicar nuestro video!❤️	@erickfancy Vayan al video original y den RT para que los chicos vean ayudenos	1
@erickfancy Vayan al video original y denle RT muchas gracias por publicar nuestro video!❤️	@erickfancy Lo mejor que he visto en mi vida 🤔🤔	1
RT @flor_vigna: Que dicen 🤔???	@flor_vigna Estas divina mi angel asi. suerte mi vida. un angel te cuida te amo @flor_vigna nuestra guerrera ..te abrazo muy fuerte	1
@flor_vigna Siiii es momento de cambiar un poco!!!! Te va a quedar divinoo!	@flor_vigna Todo te queda hermoso a vos!	1
@maduradascom @diana_dagostino Carlos Andres Perez QEPD. Dijo la verdad. Cmo estamos ahora? Jodidos y tristes x q e... https://t.co/cuB005zND5	@maduradascom @HDPY0 Eso se sabía	1
RT @AndrewsAbreu: #GUATIRE Grupo de jóvenes rescataron un espacio verde sembrando una palma bajo el elevado la gran brasa, al día siguiente...	@AndrewsAbreu Triste. Pero sembrar una palma debajo de un elevado no parece muy inteligente. Digo!	1
RT @andahazi: Cataluña acaba de regresar a 1469.	@andahazi Que angustia deben sentir los catalanes. Querido rey	1
RT @telediarior_tve: Las instituciones europeas dan la espalda a la Cataluña independiente. Tampoco la reconocen ningún país ni organización...	@telediarior_tve Es el camino que recorre cualquier país que busca su #independencia nada nuevo ni que no esté calculado	1
RT @telediarior_tve: Las instituciones europeas dan la espalda a la Cataluña independiente. Tampoco la reconocen ningún país ni organización...	@telediarior_tve Es que que país va querer apoyar a estos?? Ni la unión europea...	1
RT @MargaWirkierman: Te sorprendias con Abal Medina q estaba Duro o Duro Duro ??Mira la Gobernadora en Campaña ! Presta atencion como la mir...	@MargaWirkierman Ahí te das cuenta porque estaba con el	1
RT @luciano_eltirri: Sin palabras!! @cuervotinelli ❤️ https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli Tal cual! 🙏🙏	1
RT @luciano_eltirri: Sin palabras!! @cuervotinelli ❤️ https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli No hay palabras ❤️❤️❤️❤️🙏🙏	1
RT @luciano_eltirri: Sin palabras!! @cuervotinelli ❤️ https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli Hermoso y estoy llorando, gracias por cumplir su sueño. cada día te admiro más	1
RT @luciano_eltirri: Sin palabras!! @cuervotinelli ❤️ https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli Sin palabras,como decís. dios te bendiga siempre queriendo marcelo	1
RT @CNCOMusic: Último día para enviar sus videos... #Repost @pandicorniof * * * #mamita #mamitafanvideo 🙏🙏 https://t.co/zQ8QUiyyk	@CNCOMusic Suerte a todas 🙏❤️	1
RT @Sexy_Queens: Uno así para Navidad 🙏🙏 https://t.co/uSR52oWao	@Sexy_Queens Y si te lo regalo para el día que me des el placer de conocerte ?????🙏	1
RT @EsRecuerd0s: RT Si fuiste de la generación que se emocionaba cada vez que aparecía este msj en tu ordenador. https://t.co/enopbKRfAL	@EsRecuerd0s @ImMrMoonLight Y yo así 🙏 https://t.co/9iYNIQ07ko	1
RT @Juliolamas1: Quiero para mi País :Q se esclarezcan los casos Maldonado y Nisman.Q vayan presos TODOS los q robaron.Q se gobiernen...	@Juliolamas1 @diegobranca Esto es posible. Gracias a @mauriciomac	1
RT @Thomas_Astruc: Entonces, ¿qué piensas de este comienzo de la nueva temporada?	@Thomas_Astruc OMG Es lo mejor! GRACIAS POR TANTO PERDON POR TAN POCO	1
RT @guillermochoa: POR FAVOR: Ojalá nos puedan ayudar donando pilaquetas para mi padre, Guillermo Ochoa. Banco de sangre del hospital ABC de...	@guillermochoa Mis mejores deseos para ustedes y que se recupere tu señor padre...🙏🙏🙏🙏🙏🙏	1
RT @guillermochoa: POR FAVOR: Ojalá nos puedan ayudar donando pilaquetas para mi padre, Guillermo Ochoa. Banco de sangre del hospital ABC de...	@guillermochoa @adevaldes Ojalá pudiera mi corazón para ese gran hombre un abrazo a la distancia	1
RT @guillermochoa: POR FAVOR: Ojalá nos puedan ayudar donando pilaquetas para mi padre, Guillermo Ochoa. Banco de sangre del hospital ABC de...	@guillermochoa Ojala que se mejore pronto se le admira y se le quiere	1
RT @guillermochoa: POR FAVOR: Ojalá nos puedan ayudar donando pilaquetas para mi padre, Guillermo Ochoa. Banco de sangre del hospital ABC de...	@guillermochoa @r4ulrom4 Mis deseos de pronta recuperación a tan gran periodista. Saludos.	1
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tarde... NAKSDLENJSKSKSKS atención al siguiente tweet ❤️🙏	@MicaSuarez12 Mas RAPIDO Micaela Que me muero	1
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tarde... NAKSDLENJSKSKSKS atención al siguiente tweet ❤️🙏	@MicaSuarez12 Ya cuenta de una vez!	1
RT @yuyacst: Te amo.	@yuyacst Ya sabía 🙏🙏	1
RT @yuyacst: Te amo.	@yuyacst Y yo a ti bebe. te veo más tarde.	1
RT @yuyacst: Te amo.	@yuyacst Yo más, pero no le digas a nadie 🙏	1
RT @yuyacst: Te amo.	@yuyacst Te amo más	1
RT @yuyacst: Te amo.	@yuyacst No más que yo a ti.	1
RT @yuyacst: Te amo.	@yuyacst Yo igual	1
@guillercastello @elisacarrio Y eso que se rumoreaba de IOMA y Bara deli que onda?	@guillercastello @elisacarrio Excelente!	1
RT @EPN: México no reconocerá la declaración unilateral de independencia de Cataluña. Hacemos votos por una solución política y pacífica...	@EPN Todo México no te reconoce y tu preocupándote por España ?	1

Table A50: Examples of labeled data: Twitter (15)

RT @luciano_eltirri: Sin palabras!! @cuervotinelli ♥ https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli Sin palabras... ponerse de pie y aplausos!	1
RT @luciano_eltirri: Sin palabras!! @cuervotinelli ♥ https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli La imagen vale mas que mil palabras... un angelito de dios...	1
RT @LuisCarlos: Cada vez que dicen "toda la gente buena se fue del país", me dan ganas de armar un grupo de apoyo llamado #ElRepele.	@LuisCarlos @nelsonbocaranda "Toda la gente buena" no es preciso. "Toda la gente que pudo irse" es un término más acorde a la realidad.	1
RT @matiasmier19: Con trabajo y sacrificio, todo llega! Vamos por más, vamos @JuniorClubSA! 🏆🏆🏆 https://t.co/c4Ypf8tqR	@matiasmier19 @JuniorClubSA Felicitaciones campeón. Así se callan bocas	1
RT @LuisFlorido: Estamos en EEUU, mañana nos reuniremos con Presidente del TSJ en el exilio y magistrados legítimos ¡atentos!	@LuisFlorido Y tu pasaporte te lo dieron? Cómo?	1
RT @LuisFlorido: Estamos en EEUU, mañana nos reuniremos con Presidente del TSJ en el exilio y magistrados legítimos ¡atentos!	@LuisFlorido OK, cuanto le van a pagar al presidente por haber hablado de diálogo en la reunión que hubo en Washington la semana pasada?	1
RT @LuisFlorido: Estamos en EEUU, mañana nos reuniremos con Presidente del TSJ en el exilio y magistrados legítimos ¡atentos!	@LuisFlorido Estas de paseo y el nuevo CNE buenas noches	1
RT @LuisFlorido: Estamos en EEUU, mañana nos reuniremos con Presidente del TSJ en el exilio y magistrados legítimos ¡atentos!	@LuisFlorido Ok, estaremos atentos, pero aquí no pasa nada. Tiempo de enseñarse y administrar expectativas.	1
RT @boegob: Publicado BOE (extraordinario) num. 260, 27/10/2017.Sumarío: https://t.co/zWzkv1uZ3 #BOE	@boegob Gracias a Dios, me parece mentiría!	1
RT @boegob: Publicado BOE (extraordinario) num. 260, 27/10/2017.Sumarío: https://t.co/zWzkv1uZ3 #BOE	@boegob al fin; espero respeto al Estado de derecho y a todos los españoles.	1
RT @Elme_sero: Un héroe sin capa, sin super poderes. Un héroe de la vida real. https://t.co/qP7c9UOH2C	@Elme_sero Que bien	1
RT @Elme_sero: Un héroe sin capa, sin super poderes. Un héroe de la vida real. https://t.co/qP7c9UOH2C	@Elme_sero @ProfHoax Bien ☺ felicitaciones 🍷	1
RT @pburelli: Arrestos en Madrid deben tener MUY preocupado al jefe de todos ellos, un tal @RRamirezVE quien cree estar a salvo por ser Emb...	@pburelli @RRamirezVE @UN @ONU,es ojalá y Dios lo oiga!	1
RT @Maru_Abregu: En dietética Callao cola para comprar los tés de la familia Maldonado. Gracias trolls, ahora tenemos una forma de ayudarlo...	@Maru_Abregu Gracias por la data !!	1
RT @melortega1: Luego del compromiso demostrado durante la peor emergencia del País. GFR Media tomó la decisión de despedirme a mi y a otro...	@melortega1 Suerte, eres buena y mejores cosas vendrán. #ConLaFrenteEnAlto	1
RT @melortega1: Luego del compromiso demostrado durante la peor emergencia del País. GFR Media tomó la decisión de despedirme a mi y a otro...	@melortega1 Qué insensibles! Mejores cosas vendrán y otras puertas se abrirán	1
RT @melortega1: Luego del compromiso demostrado durante la peor emergencia del País. GFR Media tomó la decisión de despedirme a mi y a otro...	@melortega1 Exito en todo, suerte!	1
RT @basquetplus: Brillante https://t.co/sJwhOHY1dO	@basquetplus Clarísimo ! 🍷🍷🍷	1
RT @basquetplus: Brillante https://t.co/sJwhOHY1dO	@basquetplus EXCELENTE. !!!!	1
RT @basquetplus: Brillante https://t.co/sJwhOHY1dO	@basquetplus Brillante!!!! Muchos como vos Necesita este país	1
RT @dominguezja: SATURNO es la canción más importante en la radio!! Felicitades!! @pabloalboran !! Te queremos!! https://t.co/1RWXUgdiin	@dominguezja @pabloalboran Felicitades a nuestro Pablete!! Es de los te mas más bonitos que he escuchado, en toda mi Vida!! ♥♥♥🎵🎵🎵	1
RT @dominguezja: SATURNO es la canción más importante en la radio!! Felicitades!! @pabloalboran !! Te queremos!! https://t.co/1RWXUgdiin	@dominguezja @pabloalboran Felicitades a nuestro Pablete!! Es de los te mas más bonitos que he escuchado, en toda mi Vida!! ♥♥♥🎵🎵🎵	1
RT @dominguezja: SATURNO es la canción más importante en la radio!! Felicitades!! @pabloalboran !! Te queremos!! https://t.co/1RWXUgdiin	@dominguezja @pabloalboran Felicitades a nuestro Pablete!! Es de los te mas más bonitos que he escuchado, en toda mi Vida!! ♥♥♥🎵🎵🎵	1
RT @dominguezja: SATURNO es la canción más importante en la radio!! Felicitades!! @pabloalboran !! Te queremos!! https://t.co/1RWXUgdiin	@dominguezja @pabloalboran Viva Málaga y Andalucía entera Felicitades Rey te lo mereces como Nadie por fin	1
RT @dominguezja: SATURNO es la canción más importante en la radio!! Felicitades!! @pabloalboran !! Te queremos!! https://t.co/1RWXUgdiin	@dominguezja @pabloalboran Te lo mereces muchísimo, enhorabuena mi campeón ♥	1
RT @GirautaOficial: Albert Rivera: "Ahora seremos los demócratas los que gntaremos "Votarem" https://t.co/ghAmn3HcLk	@GirautaOficial Cree usted que unas elecciones a tan corto plazo es lo que interesa a Cataluña y España?	1
RT @Juanarza: Os resumo lo de Puigdemont: que no haya violencia, por favor, porque entonces me caen 30 años seguro.	@Juanarza Exacto, fue la verdadera intención de su discurso hoy	1
@lucyanaah: Pocos días falta para ver qué pasará con nuestro . No me cabe duda que la gran mayoría quiere lo mismo, oportunidades, tr...	@lucyanaah @Zurdos_No_More Me gustan sus deseos . La realidad es que no hay ningún candidato capaz de hacerlo .	1
RT @SilvestreFDC: Al final me quedo con las mujeres q les gustó, no con las q me gustan! Triste terminar de hacer el amor y no tener de que...	@SilvestreFDC Eso le pasa por no ser fiel..	1
RT @perezreverte: @corrupcoencien Usar "cesar" como se usa estos días es una incorrección. Nadie puede cesar a nadie. Se destituye a alguien...	@perezreverte @corrupcoencien Bueno tienes razón,peroNo damos para mas.Este es el nivel.	1
@alejandrosanmo: Este chico ha pedido por favor que no se difundan ni su nombre ni sus fotos. Se llama Antonio Manuel Guerrero Escudero...	@alejandrosanmo Que pena se me ha escapado el dedo...	1
RT @marley_ok: Gracias!!!! ♥♥♥♥ https://t.co/wCs497JUnP	@marley_ok Qué alegría!	1
RT @marley_ok: Gracias!!!! ♥♥♥♥ https://t.co/wCs497JUnP	@marley_ok Que felicidad	1
RT @DValdebenitoA: Estimados evangélicos: Sabían que si marcan 2 opciones (Piñera y Kast) y escriben CV (Cristo Vive) se deroga ley de abor...	@DValdebenitoA gracias por el dato, eres muy inteligente.	1
@esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @DeniseDresserG Claro!! Ya lo amenazaron!!	1
@mcg1465: A esta perrita sus dueños NO la quieren más! SIN PALABRAS Esta castrada, la trasladan donde alquien le de un lugarcito.Est...	@mcg1465 Su mirada lo dice todo...	1
@carnecrudaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? Pública...	@carnecrudaradio Aquí.no hay,lo que tiene que haber !!!!	1
@PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de iasRenovables ht...	@PRODUCCION_ARG Brillante!!!	1
@PRODUCCION_ARG: Hoy se firmó un acuerdo entre empresas, sindicatos y el gobierno para aumentar la producción de iasRenovables ht...	@PRODUCCION_ARG Excelente !!!!	1
@BibiMonteroDR: Cómo es posible que en el extranjero sí hablen de la corrupción en México y aquí las malditas televisoras callan! ...	@BibiMonteroDR @GarcaMario Es que la gente se la pasa viendo telenovelas y les embrutece el cerebro.	1

Table A51: Examples of labeled data: Twitter (16)

@Saul_Craviotto: Creo que ha llegado la hora de pronunciarme. No me podía callar más. ña	@Saul_Craviotto @policia Estamos muy orgullosos de ti tanto por tus tronc os deportivos como por tu arrojo personal un abrazo desde zaragoza	1
@Saul_Craviotto: Creo que ha llegado la hora de pronunciarme. No me podía callar más. ña	@Saul_Craviotto Estamos muchísimos millones de españoles con vosotros. Un millón de gracias...	1
@UniCiudadanaAR: Máximo: "Pónganse de pie, no bajen la vara. Hay futuro si ustedes no se acostumbran". http...	@UniCiudadanaAR @mayrasmendoza me parece a mi o son 5 gatos locos? Pensar que iban por todo, jaja	1
@RubiU5: Youtube va un poco lento hoy No se procesa bien el video	@RubiU5 Yo ya lo vi y no paso nada	1
@SF9official: ¿Esto es todo?	@SF9official ¿Que quieres decir con esto es Todo? Dame más pistas. ¿Como podes ser tan lindo?	1
@bysTaXx: Aquí con @ByRubi9	@bysTaXx @ByRubi9 ME ENCANTA	1
@AnonymousVene10: Que memoria tan corta tenemos los Venezolanos cada quien siguió con su vida y olvidó a los q ya no tienen la bendición...	@AnonymousVene10 @untal_no No es así, NO manipules!!	1
@bryanmouque: Todos queremos libertad.	@bryanmouque Todos tenemos derecho a ser libres	1
@astros: LA PIÑA!	@astros Viva La Piña!!!!	1
@jtgazmuri: ¿O los ASESORES que ayudaron a la Josefa a desfondar el Municipio de Providencia?	@jtgazmuri Que manera de hablar sandeces por parte de se entera ni por la prensa.	1
@susana_zori: Hoy hace un año q falleció mi hijo Pablo atropellado y abandonado por un borracho al volante. Ahora quiero cárcel para él...	@susana_zori @UsinadeJusticia Lo siento mucho	1
@susana_zori: Hoy hace un año q falleció mi hijo Pablo atropellado y abandonado por un borracho al volante. Ahora quiero cárcel para él...	@susana_zori @UsinadeJusticia Que terrible Susana, te admiro la fuerza.	1
@PiensaPrensa: [LO ULTIMO] PIÑERA: "Si la Presidenta cierra Punta Peuco, voy a volver a abrirla".	@PiensaPrensa Por qué se les debe un trato especial? Son asesinos y deben estar en cárceles comunes con criminales...	1
@zibenial: Casi me echan del bar de la risa	@zibenial lo peor que no estaba mirando	1
@zibenial: Casi me echan del bar de la risa	@zibenial Jaja No lo había visto	1
@zibenial: Casi me echan del bar de la risa	@zibenial Es tremendo, el tipo sin querer y de la nada, hizo el mejor spot de campaña de su vida.	1
@christophervele: 4 4 son el mejor fandom del mundo!	@christophervele Te amo ! Me encantan !	1
@christophervele: 4 4 son el mejor fandom del mundo!	@christophervele Gracias les queremos y daríamos todo por ustedes	1
@dvlzura: si leiste "una nota gratis".	@dvlzura tal vez tenga mala la vista	1
@eltanomdp: No olvidar a estos mercenarios oficialistas son cómplices de una desaparición forzada seguida de asesinato htt...	@eltanomdp Más rompe pelota están con esto..	1
@GonzGravano: Y YouTube es cuestión de tiempo, entiéndanme que quiero hacer algo distinto y no el contenido que hacen todos. Paciencia...	@GonzGravano Yo te estoy dando más visitas porque con los que tenes los repito mil veces	1
@cncxmendes: Me muerdo de la vergüenza pero acá les dejo mi cover de -	@cncxmendes @CNCOMusic Buah... a penas escuche tu voz me recordaste a Laura Esquivel. Muy bien cantado.	1
@cncxmendes: Me muerdo de la vergüenza pero acá les dejo mi cover de -	@cncxmendes @CNCOMusic Me encanta	1
@cncxmendes: Me muerdo de la vergüenza pero acá les dejo mi cover de -	@cncxmendes @CNCOMusic Que bonito cantas deberías hacer cover de todas las canciones	1
@annapacheco_: Hoy empieza el juicio a La Manada y hace un año escribi esto sobre grupos de WhatsApp. Igual cuesta admitirlo, pero algu...	@annapacheco_ @MrPINTXO No lo conocía y me ha parecido súper interesante. Gracias a @annapacheco_ por escribirlo..	1
@erickfancy Vayan al video original y denle muchas gracias por publicar nuestro video!	@erickfancy VAYAN AL VIDEO ORIGINAL Y DEN PARA QUE LOS CHICOS LOS VEAN AYUDENOS	1
@Thomas_Astruc: Entonces, ¿qué piensas de este comienzo de la nueva temporada?	@Thomas_Astruc Wow qué increíble y fabulosos	1
@EPN: México no reconocerá la declaración unilateral de independencia de Cataluña. Hacemos votos por una solución política y pacífica.	@EPN Muy seguramente no conoces nada del tema para variar y das tu punto de vista, así han sido 5 años de tu triste y equívoca presidencia	1
@CGurisattiNTN24: Una cosa es un campesino con una hectárea o menos de coca bajo el dominio de un narco violento y otra es un narco disf...	@CGurisattiNTN24 Da lo mismo, a quien terminan vendiéndole esa COCA a los llamados pequeños cultivadores de coca ?...	1
@matiasmier19: Con trabajo y sacrificio, todo llega! Vamos por más, vamos @JuniorClubSA!	@matiasmier19 @JuniorClubSA Vamos por todo!	1
@boegob: Publicado BOE (extraordinario) num. 260, 27 10 2017.Sumarior:	@boegob Gracias a Dios, me parece mentira!	1
@Elme_sero: Un héroe sin capa, sin super poderes. Un héroe de la vida real.	@Elme_sero Que bien	1
@SilvestreFDC: Al final me quedo con las mujeres q les gustó, no con las q me gustan! Triste terminar de hacer el amor y no tener de que...	@SilvestreFDC papasito como me encantas :? :? :? hablemos tu y yo lol :? mi unico idolo vallenato... i love so m...	1
@SilvestreFDC: Al final me quedo con las mujeres q les gustó, no con las q me gustan! Triste terminar de hacer el amor y no tener de que...	@SilvestreFDC "Mujeres?" en plural? Luego no es casado? creo que a las mujeres en general nos gusta oír	1
@FrayJosepho: Ni queman los DNIs y los pasaportes, ni Rufián renuncia a su sueldo de diputado ni el Bar abandona la Liga. ¡Vaya mierda...	@FrayJosepho @ManuelBallbe jajajaja!!! me encanta lo del Bar a!!!!peor que los niños pequeños en un divorcio	1
@perezreverte: @corrupcoencien Usar "cesar" como se usa estos días es una incorrección. Nadie puede cesar a nadie. Se destituye a alguien...	@perezreverte @corrupcoencien Bueno tienes razón, pero ...No damos para mas. Este es el nivel.	1
@onielluipig: Pues he estado estudiando el caso un poco y creo que sé lo que pasó ayer con la... Lo explico aquí	@onielluipig La verdad una visión fría y racional después del esperpento de ayer es de agradecer! Ojalá sea así...	1
@dianamotavo: Acabo de ver el especial @noticiasrcn defendiendo a Santiago Uribe y recorde cuando decían q Gaseosas eran menos perjudiciales...	@dianamotavo @NoticiasRCN Se nota bastante la mano detrás de estas marionetas que se hacen llamar periodistas...	1
@protestona1: Curioso que estén en la cárcel dos Jordis por manifestarse y ande suelto el que ha expoliado Cataluña durante más de 30 añ...	@protestona1 @agariaportela Y el instigador en la sombra del nacionalismo catalán.....	1
@marley_ok: Gracias!!!!	@marley_ok Qué alegría!	1

Table A52: Examples of labeled data: Twitter (17)

@DValdebenitoA: Estimados evangélicos: Sabían que si marcan 2 opciones Piñera y Kast) y escriben CV (Cristo Vive) se deroga ley de abor...	@DValdebenitoA y esta bien vivo y murió para que no fueras a la muerte eterna, aunque seas muy fome, igual lo hizo	1
@DValdebenitoA: Estimados evangélicos: Sabían que si marcan 2 opciones Piñera y Kast) y escriben CV (Cristo Vive) se deroga ley de abor...	@DValdebenitoA gracias por el dato, eres muy inteligente.	1
@ziberia1: Casi me echan del bar de la risa	@ziberia1 No puedo parar de reirme!!!!	1
@AngeldebritoOk: La mejor noche de @MicaViconte	@AngeldebritoOk @MicaViconte La mejor noche la rompieron	1
@AngeldebritoOk: La mejor noche de @MicaViconte	@AngeldebritoOk @MicaViconte Excelentes sin duda los nuevos campeones del bailando 2017	1
@JosepLago: Hola, @pnique soy un rojo y coordino las juventudes de @Societatcc. Mañana estaré organizando la manifestación junto a amigo...	@JosepLago @pnique @Societatcc Yo soy de derechas pero me quito el sombrero ante ti	1
@luisafemandaw: El gusto y el amor son cosas completamente diferentes.	@luisafemandaw Obviamente, tu AMAS a Andres y te gusta un poquito simón	1
@pnique: Si la manifestación la convocan los que odian a los gays, a los inmigrantes, a los rojos y a los catalanes, a lo mejor para uni...	@pnique Manipular a la gente diciendo que la manifestación es contra los catalanes no te dará más votos.	1
@dosogas: A quien le haremos la videollamada la próxima semana? (De todas las redes sociales, los más activos)	@dosogas Ojala ami los amo me la paso despierta para ver sus videos y cuando dijieron videollamada grite	1
@dosogas: A quien le haremos la videollamada la próxima semana? (De todas las redes sociales, los más activos)	@dosogas Soy re yeta a mi seguro que no, pero me encantaria aaaaaa	1
@GirautaOficial: Es un momento muy oportuno para ver este video con mucha atención.	@GirautaOficial Increible.	1
@AlbanoDante76: "Independentistas golpean a un chico que no quiso gritar 'Visca Catalunya independent'". Imagináis las portadas? Pues fu...	@AlbanoDante76 el problema de dirección de Podemos lo tienes muy bien identificado, tienes razón!	1
@ehbildu: Tenemos que construir un muro democrático entre todos los democratas @ArnaldoOtegi	@ehbildu @ArnaldoOtegi Vosotros el único muro que podéis construir es el muro de la vergüenza, amigos de asesinos,nazis,Ratas sarnosas	1
@AromaDeJesus: Descansa en Dios, mañana la gloria del Señor se manifestará en tu vida Confía, ten Fe y Cree en Él.	@AromaDeJesus Amen mi señor siempre a tu lado con todo mi corazón	1
@IgorMeltxor: Sin ellos nada de esto sería posible. No lo olvidéis nunca.	@IgorMeltxor @kanallak1 Ni olvido ni perdon	1
@alejandrosanmo: ¿Os acordáis del acoso en Atocha a politicos catalanes? Mirad quién era uno de ellos.	@alejandrosanmo Si es él, era el que denunciaba "los atropellos sufridos desde que Carmena 'asaltó el cielo'" en El Español.	1
@Ferliglesias: La jubilación sin aportes bancada por los demás jubilados es lo que hizo que 75% de los jubilados terminaran cobrando la m...	@Ferliglesias Van a saquear a los jubilados. La rueda gira, y en unos años te irás que dar explicaciones	1
@Ferliglesias: La jubilación sin aportes bancada por los demás jubilados es lo que hizo que 75% de los jubilados terminaran cobrando la m...	@Ferliglesias @pulsionparcial La jubilación sin aportes es cuestionable... Los servicios domésticos no tenían aporte...	1
@DaniMateoAgain: Un poquito de polémica de buena mañana? Venga... O pinad!	@DaniMateoAgain Más claro no se puede decir.La primera puta norma de la democracia es el derecho a decidir.	1
@Sprya_: Me parece un crimen sinceramente	@Sprya_ También me parece un crimen, que pelo tan hermoso tenía	1
@VMatterfilm: No cuesta nada y les tocara las narices porque no podrán romperlos	@VMatterfilm UAAAAU... Claro, conseguirás mucho con eso... no se cómo no lo han pensado los de Anonymous	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Me encantaria santo padre que se acuerde de todos los argentinos no solo a los delincuentes que estafaron al país	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Con Dios todo sin el nada nada somos iluminados Jesucristo para llevar tu palabra a nuestros hermanos	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Acaso no creemos que Jesus murio y resucito? Asi tambien Dios resucitara con Jesus a los que han muerto...	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Aleluya! El señor resucito! Hoy es el día glorioso di señor! Sea mos testigos.vamos a misa a darle gracias...	1
@PerdigueroSIPep: Otro Mosso que no merece llevar el uniforme que lleva por sectario y amenazar a un ciudadano que solo exhibe la bandera...	@PerdigueroSIPep Si se autoasca usted mismo. Le dice "si vuelve a levantar la mano así, le denuncio"	1
@PerdigueroSIPep: Otro Mosso que no merece llevar el uniforme que lleva por sectario y amenazar a un ciudadano que solo exhibe la bandera...	@PerdigueroSIPep @fororc De esto seguro que no sabes nada, a qué no?	1
@joan_guirado_: Una discoteca de ofrece barra libre gratis esta noche, para celebrar la orden de detención de @KRLS Puigdemont...	@joan_guirado_ @KRLS Cuando será la celebración de la detención de los de la Gurtel?	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Amén por su sacrificio fuimos redimidos...ya no vivo yo más Cristo vive en mí!	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Si CRISTO, Es EL SEÑOR DE TU VIDA, Tienes Que Vivir Esperanzado, Porque Ahora LA ESPERANZA, Vive En Ti.	1
@ramoncotarelo: La única violencia en Cataluña la traen los fascistas españoles amparados x el gob o a sueldo suyo x la unidad N https://...	@ramoncotarelo Y vosotros santos! Si os esperarás un poquito o hago un altar! Gracias por las pintadas en la puerta...	1
@leotaxil: Hay que replantearse si tienen sentido las monarquías hoy en día ¿Sabéis lo que le cuesta al Estado la familia Borbón? https://...	@leotaxil Si, a cada catalan la monarquia le cuesta 0,17 euros, o sea 17 centimos y a cada catalan la TV3 le cuesta 30 euros al año.	1
@leotaxil: Hay que replantearse si tienen sentido las monarquías hoy en día ¿Sabéis lo que le cuesta al Estado la familia Borbón? https://...	@leotaxil Más o menos va así la cosa.	1
@Profeta_Baruc: El hilo de hoy comienza con una pregunta: ¿Qué tienen en común estas tres fotos?	@Profeta_Baruc @kezeyo Lo mantenía y mantiene el pueblo	1
@Profeta_Baruc: El hilo de hoy comienza con una pregunta: ¿Qué tienen en común estas tres fotos?	@Profeta_Baruc Muchas gracias por la historia y la distracción tan necesaria en este momento	1
@elmundoes: Jueces progresistas arremeten contra la juez Lamela por su decisión de encarcelar a ocho ex 'consellers'.	@elmundoes Y que hubieran hecho ante el desacato y violacion de todas las leyes que se han saltado	1
@elmundoes: Jueces progresistas arremeten contra la juez Lamela por su decisión de encarcelar a ocho ex 'consellers'.	@elmundoes Al pueblo nos da miedo que los jueces se critiquen entre sí sus decisiones.	1
@elmundoes: Jueces progresistas arremeten contra la juez Lamela por su decisión de encarcelar a ocho ex 'consellers'.	@elmundoes Estos jueces para la progresia no consideran acaso como excepcional y grave	1
@elmundoes: Jueces progresistas arremeten contra la juez Lamela por su decisión de encarcelar a ocho ex 'consellers'.	@elmundoes @MainatJM Osea que se constata que la jueza no es progresista	1
@MariaCorinaYA: Llámelo como lo llame. Maduro anunció la bancarrota de Venezuela. Ahora, ya nadie lo duda, la única opción es sacarlos d...	@MariaCorinaYA Y como lo sacamos??Cuál es la estrategia y acciones?? N o siga diciendo cosas que no puede cumplir	1

Table A53: Examples of labeled data: Twitter (18)

@MariaCorinaYA: Llámelo como lo llame, Maduro anunció la bancarrota de Venezuela. Ahora, ya nadie lo duda, la única opción es sacarlos d...	@MariaCorinaYA Noooooooooo, de veras?????	1
@MariaCorinaYA: Llámelo como lo llame, Maduro anunció la bancarrota de Venezuela. Ahora, ya nadie lo duda, la única opción es sacarlos d...	@MariaCorinaYA Bueno, se la pongo fácil empiece por el tachira a la calle y el país se prende rápido yo lo se lo seguro	1
@MariaCorinaYA: Llámelo como lo llame, Maduro anunció la bancarrota de Venezuela. Ahora, ya nadie lo duda, la única opción es sacarlos d...	@MariaCorinaYA Pero solo hablando no lo vamos a sacar hay que actuar in mediatamente llamar al pueblo a la calle	1
@MariaCorinaYA: Llámelo como lo llame, Maduro anunció la bancarrota de Venezuela. Ahora, ya nadie lo duda, la única opción es sacarlos d...	@MariaCorinaYA Eres la única que tiene bien claro que se debe hacer y que se debe exigir para obtener la victoria y levantamiento del país.	1
@_ju1_: Tras acabar una marcha española en Mataró. "Di viva España o t e matamos". El agredido, hospitalizado. Difunde	@_ju1_ Cuando dice lo de Viva España? No lo oigo...	1
@pnique: "¿Tú eres diputada?" Para ser andalza, eres más educada que yo "	@pnique Yo creo que no, eso solo lo hacen hombres sin escrúpulos, acompañados, y cobardes	1
@pnique: "¿Tú eres diputada?" Para ser andalza, eres más educada que yo "	@pnique Ahhhhh y no entres en el tema machismo, que ella se sabía defender bien.	1
@jlespert: Sisiti mi tesoro. Toooda la flexibilidad laboral que quieras, pero también quiero a Amazon acá mordiéndote los talones, sino...	@jlespert Ellos son un monopolio que hace lo que quiere con los usuarios y sobre todo cuando cobras con mercado pago y trabajan con tu plata	1
@jlespert: Sisiti mi tesoro. Toooda la flexibilidad laboral que quieras, pero también quiero a Amazon acá mordiéndote los talones, sino...	@jlespert Gracias pelado, sos la voz de los estafados por impuestos y los esnartados por los monopolios!.. gracias	1
@VMatterfilm: Me preguntan compañeros de la prensa holandeses si realmente Catalunya puede organizar una marcha con miles y miles de coc...	@VMatterfilm Si para una final de champions nos movemos calcula para algo importante.	1
@VMatterfilm: Me preguntan compañeros de la prensa holandeses si realmente Catalunya puede organizar una marcha con miles y miles de coc...	@VMatterfilm Poder, podemos. Ya está en marcha.	1
@VMatterfilm: Me preguntan compañeros de la prensa holandeses si realmente Catalunya puede organizar una marcha con miles y miles de coc...	@VMatterfilm @MainatM Diles que sí!!!	1
@VMatterfilm: Me preguntan compañeros de la prensa holandeses si realmente Catalunya puede organizar una marcha con miles y miles de coc...	@VMatterfilm @JuntPeSi_Cat No son ni la mitad de los catalanes y mira el follón con su migración!	1
@perezfecto: Viernes existenciales.	@perezfecto Jajajajaja buena	1
@Argengchogau: AYER SÁBADO por la tarde Río Gallegos.CFK saliendo por puerta de servicio. DESCUBIÉ A. vecinos le gritan chorra, ladro...	@Argengchogau Era ella!ahora va y antes no pudo ni siquiera ir a votar la democrática?cinica corrupta.chorra, delincuente!!!	1
@Argengchogau: AYER SÁBADO por la tarde Río Gallegos.CFK saliendo por puerta de servicio. DESCUBIÉ A. vecinos le gritan chorra, ladro...	@Argengchogau @alitwitea Tampoco gran escarcha...Dos minas y un célula, el tema es q no puede moverse libremente	1
@iamjulianwilbur: @jordievole Atentado de La Rambla: semanas en shock A tentado en NYC: pasamos página en horas.Atentado en Somalia: n...	@iamjulianwilbur @jordievole Esa misma reflexión vale para los de USA o C anadá??.. o para los de Japón?? o para los d Egipto??...	1
@GLlamazares: Soy tan adversario político del PP como del independentismo. Las políticas que defienden perjudican a la democracia y a lo...	@GLlamazares Porque los nacionalistas españoles no se consideran ellos mismos nacionalistas?	1
@MisOdios: Los que pedían que Bullrich se tire por la ventana por las mentiras de un mapuche ahora piden que no se prejuzgue a Boudou. Q...	@MisOdios @vagge, delia Son los mismos a los que no se les movió un pelo cuando vieron descalzo a un alumno abandonado.	1
@Silvialbarzabal: SE DIERON CUENTA QUE EL TERRORISMO ISLÁMICO SE LEVÓ LA VIDA DE 5 COMPATRIOTAS Y NINGUNA ORGA DE DDHH MANIFESTÓ SU TR...	@Silvialbarzabal @PaisPPT Hay que salir a enfrentarlos pero no uno solo sino en grupo uno solo no hace nada	1
@Silvialbarzabal: SE DIERON CUENTA QUE EL TERRORISMO ISLÁMICO SE LEVÓ LA VIDA DE 5 COMPATRIOTAS Y NINGUNA ORGA DE DDHH MANIFESTÓ SU TR...	@Silvialbarzabal @PaisPPT Si. Me di cuenta. Pero este país es un chotada.	1
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 @NunkMasKK No les sale ni una encuesta jajaja	1
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 Entre los que laburan y pagan impuestos, y los vagos que quieren todo de arriba.	1
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 Siempre fue pago.. Antes lo pagábamos todos, y con sobreprecios	1
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 SIEMPRE SE PAGO, antes lo hacía el estado con la plata de todos, ahora lo paga solo el que lo quiere ver, me parece justo	1
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 quiero ver a los dueños de Indalo y Radio 10 poner la \$ que se robaron de Estado.	1
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 Creo que el futbol sería gratis	1
@danielsenderos: Si en España hubiera tanta fotovoltaica como en Alemania, la electricidad nos costaría un 36% menos	@danielsenderos @marinu77 Con las placas el usuario en todo el mundo puede vender lo que no gaste.	1
@danielsenderos: Si en España hubiera tanta fotovoltaica como en Alemania, la electricidad nos costaría un 36% menos	@danielsenderos @marinu77 Los políticos aquí no deja ponerlas.	1
@danielsenderos: Si en España hubiera tanta fotovoltaica como en Alemania, la electricidad nos costaría un 36% menos	@danielsenderos ya harían algo para que nos costara igual o más	1
@danielsenderos: Si en España hubiera tanta fotovoltaica como en Alemania, la electricidad nos costaría un 36% menos	@danielsenderos @desdefuer Yo creo que las eléctricas ganarían un 36 por ciento más.	1
@danielsenderos: Si en España hubiera tanta fotovoltaica como en Alemania, la electricidad nos costaría un 36% menos	@danielsenderos Costaría lo mismo, ganaría el Estado en impuestos para mejorar las condiciones de vida	1
@Pablangasss: Que? En 2 meses se termina el año? Cuando pasó todo esto ????	@Pablangasss Ah pasado tan rápido todo ahre	1
@Pablangasss: Que? En 2 meses se termina el año? Cuando pasó todo esto ????	@Pablangasss De todo algunos hasta nos cambio la vida	1
@ibepacheco: La indefensión y desamparo de los periodistas se evidencia cuando un colega desaparece después de amenazas y no hay autoridad...	@ibepacheco Y la sociedad de periodistas que no están ahí para eso.	1
@ibepacheco: La indefensión y desamparo de los periodistas se evidencia cuando un colega desaparece después de amenazas y no hay autoridad...	@ibepacheco la misma indefensión que vive el ciudadano común! Por eso el venezolano huye dejando aca su corazón	1
@ibepacheco: La indefensión y desamparo de los periodistas se evidencia cuando un colega desaparece después de amenazas y no hay autoridades	@ibepacheco Cuál autoridades? Acá NO hay gobierno si no un grupo de vagos Mail entretenidos que asaltaron el poder...	1
@xavialdekoa: Tocando el cielo con los dedos. Una fotografía maravillosa	@xavialdekoa Maravilla de foto.	1
@xavialdekoa: Tocando el cielo con los dedos. Una fotografía maravillosa	@xavialdekoa Qué maravilla	1
@xavialdekoa: Tocando el cielo con los dedos. Una fotografía maravillosa	@xavialdekoa ¡Qué maravilla de foto!	1

Table A54: Examples of labeled data: Twitter (19)

@AlbertoBernalle: Trataré de ser lo más sutil posible: si a usted le parece bien q alias @TimoFARC sea candidato presidencial, usted es...	@AlbertoBernalle @TimoFARC Exacto	1
@PaulaVazquezTV: Y mientras sacudes tu bandera	@PaulaVazquezTV y la vida sigue su curso...	1
@PaulaVazquezTV: Y mientras sacudes tu bandera	@PaulaVazquezTV @pnique Todo eso está muy bien. Pero no creo que Rajoy convenciera a los separatistas que proclamarán la independencia. A	1
@680carlos: Las Universidades huyen del pensamiento critico. Los alumnos son clientes no estudiantes. Es el negocio de la mediocridad.	@680carlos Y los colegios también...	1
@UltimaHora_Info Entonces el PP y Ciudadanos deberían estar prohibidos porque sólo éstos consiguen votos cuando se...	@UltimaHora_Info Por qué en España no? Yo quiero votar eso.	1
@UltimaHora_Info Entonces el PP y Ciudadanos deberían estar prohibidos porque sólo éstos consiguen votos cuando se...	@UltimaHora_Info Pero en España no	1
@VidalQuadras: Venezuela está en colapso económico total. No puede pagar su deuda y la inflación es del 2000%. Los votantes de Podemos d...	@VidalQuadras @PorteroTriano Con la que tiene Maduro encima todavía se permite el lujo de echar porquería a España. Ya le vale!	1
@VidalQuadras: Venezuela está en colapso económico total. No puede pagar su deuda y la inflación es del 2000%. Los votantes de Podemos d...	@VidalQuadras A ellos les da igual, piensan en primera persona y en vivir de subvenciones, al resto del país...ya si acaso....	1
@Lautiroman1995: Facundo Moyano critica la reforma laboral porque esta involucrados los intereses sindicales. No le importa los derechos...	@Lautiroman1995 Facundo moyano en su puta vida laburó.	1
@MainatJM: ¡A ver si ahora va a ser delito cagarse en quien te apetezca!	@MainatJM Pero es que no saben hacer nada más que poner querrelas cuando alguien dice lo que opina aunque no guste? No tienen palabras?	1
@xavialdekoa Espectacular fotografía.	@xavialdekoa Tocando el cielo con los dedos. Una fotografía maravillosa	1
@xavialdekoa Espectacular fotografía.	@xavialdekoa Maravilla de foto.	1
@xavialdekoa Espectacular fotografía.	@xavialdekoa Qué maravilla	1
@xavialdekoa Espectacular fotografía.	@xavialdekoa ¡Qué maravilla de foto!	1
@xavialdekoa Espectacular fotografía.	@xavialdekoa @Ramon_Perello Peazo FOTÓN!	1
@xavialdekoa Espectacular fotografía.	@xavialdekoa preciosa	1
@YoSoyJustin_: Lo de Lacava en Valencia fue un Show. Durante su campaña visitó ese galpón y no dijo nada (VIDEO)	@YoSoyJustin_ Fue una venganza. No dijo nada porque no estaba montado en el coroto, ahora es distinto	1
@YoSoyJustin_: Lo de Lacava en Valencia fue un Show. Durante su campaña visitó ese galpón y no dijo nada (VIDEO)	@YoSoyJustin_ Todo es un SHOW, ellos tienen el monopolio de la alimentación, el régimen son los Bachaqueros	1
@gabrielHLucero: Valores de estética visual y estética moral	@gabrielHLucero Creo que es el mejor. Estética y moralmente, el mejor de todos.	1
@gabrielHLucero: Valores de estética visual y estética moral	@gabrielHLucero Que ganas irrefrenables de bajarle la tapa de la cama solar	1
@gabrielHLucero: Valores de estética visual y estética moral	@gabrielHLucero El cansancio moral te lo bancas por resignación; pero el cansancio estético cómo haces? Es todo taan "cache"!	1
@auronplay: En 2 horas es mi cumpleaños, espero que me llegue una tarjeta de felicitación POR PA E DE UNA EMPRESA QUE NO QUIERO NOMBRAR...	@auronplay Feliz cumpleaños	1
@auronplay: En 2 horas es mi cumpleaños, espero que me llegue una tarjeta de felicitación POR PA E DE UNA EMPRESA QUE NO QUIERO NOMBRAR...	@auronplay te doy mis felicitaciones en nombre de toda la empresa	1
@Samsung_Peru: Porque creemos en ti	@Samsung_Peru vamos paolo todos estamos contigo vamos peru	1
@Samsung_Peru: Porque creemos en ti	@Samsung_Peru Yo no!!	1
@Samsung_Peru: Porque creemos en ti	@Samsung_Peru Porque creo en ti VAMOS VAMOS PERU	1
- Seguir a @RainbowEsp	@RainbowEsp A probar suerte	1
@RubiusU: He puesto una canción que me encantaba como alarma del despertador y he acabado detestándola. No cometáis el mismo error que yo...	@RubiusU Me Paso exactamente lo mismo	1
@PPTeamRafa: Si supieran todo lo que nos pasa antes de abordar un vuelo, podríamos hacer películas de eso. Nunca tenemos un viaje tranquilo...	@PPTeamRafa A mi no, la suerte me acompaña	1
@MorenoG_Agustin: No hay mayor cobardía que la de Felipe González: 14 años en el poder y con mayorías absolutas y no hizo NADA por las v...	@MorenoG_Agustin @IndEnCastellano Si que hizo sí!	1
@BeatrizTalegon: He venido en el tren con unos cuantos piolines. Constató que algunos, por sus vestimentas, venían a infiltrarse y a lia...	@BeatrizTalegon Eres parte, no te creo	1
@PipeHenriquezO: Que importante lo que me dice este Bot. Por eso este próximo 19 de noviembre, hay que levantarse a votar, para BOTAR al...	@PipeHenriquezO Para su "trabajo" de bot le basta el slogan de Piñera vienen tiempos mejores bla bla bla	1
@_anapastor_: La fiscalía belga respondiendo a todas las preguntas... ¡guay que aquí!!!	@_anapastor_ A ver, no son belgas!!!!	1
@ForreMarti: La sexta noche es el Salvame de la política.	@ForreMarti @FAQSTV3 @BeatrizTalegon Todo depende del nivel del público al que va dirigido	1
@SoyAmargada por favor esa carita!! me muero por Dios, gracias a toda esa gente que no les entra el corazón en el...	@SoyAmargada Se me partió el alma.	1
@SoyAmargada por favor esa carita!! me muero por Dios, gracias a toda esa gente que no les entra el corazón en el...	@SoyAmargada Con tanto sufrimiento gracias a Dios vive gracias a los que le salvaron bendiciones.	1
@ManuelTurizo_: Hay días en los que te pienso más que nunca	@ManuelTurizo_ Igualmente bebe	1
@mariobautista_: No se esperaban ese 11:14	@mariobautista_ Ese no,pero lo esperaba	1
@josejacas: le pasa algo a mis animoj	@josejacas Espectacular. Esto tiene futuro	1
@pabloharour: Cuando los independentistas se enteren de que los miembros de la Fiscalía de Bruselas los nombra el Rey belga van a flipar.	@pabloharour @Enlaparada Claro que sí	1

Table A55: Examples of labeled data: Twitter (20)

@plan353: Qué Soledad tan verraca la de @AlvaroUribeVel .Con la ayuda de Dios, este señor será quien saque al País de semejante atollad...	@plan353 @AlvaroUribeVel Dios nos libre	1
@brozoxmiswebs: ASF REPO A: PEMEX GASTO 114 MELONES EN DRONES PA UBICAR HUACHICOLEO, JAMÁS LOS USARON.	@brozoxmiswebs Obvio!! Y se puede saber donde están los drones? Quien iba a pilotar los drones?	1
@DaniMateoAgain: La oposición debería preguntarse seriamente qué está haciendo mal cuando no gana ni con esto.	@DaniMateoAgain La tengo! Ustedes tienen los Medios y el pueblo ha sido desposeído. Sólo nos quedan las redes sociales.	1
@DaniMateoAgain: La oposición debería preguntarse seriamente qué está haciendo mal cuando no gana ni con esto.	@DaniMateoAgain @democraciareal Oposición?? Qué oposición? Yo solo veo cómplices...	1
@DaniMateoAgain: La oposición debería preguntarse seriamente qué está haciendo mal cuando no gana ni con esto.	@DaniMateoAgain @ramonlobo También, como país, nos lo deberíamos preguntar. ¿Toleramos la corrupción y o toleramos...	1
@DaniMateoAgain: La oposición debería preguntarse seriamente qué está haciendo mal cuando no gana ni con esto.	@DaniMateoAgain @ramonlobo A la oposición deberíamos preguntarle seriamente que está escondiendo cuando no gana ni con esto.	1
@atajoslargos: En el hotel, pedi pasta de dientes. No sé si usarla o cuidarla y verla crecer y apoyarla en las decisiones que tome en su...	@atajoslargos @Anamatxu Merece un hogar donde crecer feliz.	1
@Juanjaramilloe Ya desperté y quiero que sepas que te amo mucho bebé	@Juanjaramilloe Te amooooo inexplicablemente bebe un beso de Perú hasta donde quiera que vallas hermoso. QUE TE V...	1
@Juanjaramilloe Ya desperté y quiero que sepas que te amo mucho bebé	@Juanjaramilloe Acabó de verlo, yo te amo mucho más mi bebé elefante	1
@Michaelronda: @_ruggero te amo cabron ! Eres un campeón bro !	@Michaelronda @_ruggero Los amo	1
@MaiteOficial: Q tengan un buen dia todos	@MaiteOficial Igualmente hermosa	1
@Michaelronda: @_ruggero te amo cabron ! Eres un campeón bro !	@Michaelronda @_ruggero Los amo sois los mejores	1
@Juandedios_P: Descansa preciosa	@Juandedios_P Igual bebé	1
@AlbanoDante76: Tertuliana de @la5extaTV diciendo que la vida en la arcel es "amable" con "siestas de dos horas". Cuanta miseria.	@AlbanoDante76 @luisllach @la5extaTV Pues, se podría ir ella a pasar una temporada para hacer siestecitas.	1
@alejandrosanmo: ¿Os acordáis del acoso en Atocha a políticos catalanes? Mirad quién era uno de ellos.	@alejandrosanmo @MNANoticias @malditobulo Decídmelo qué no es cierto.	1
@villalobossebas: Buenos días, acabo de despertar y soñé que corría en un campo abierto lleno de piezas del cubo. Varios de ustedes están...	@villalobossebas Yo paso 16 horas pensando en ti y 8 soñando contigo	1
@sebastianpinera: Pobre y triste espectáculo dan candidatos q, en vez de proponer ideas y soluciones para Chile, se descalifican y sacan...	@sebastianpinera Señor usted es un androide estoy seguro.	1
@AbelPintos: ¡BUENAS!	@AbelPintos Lindo, Siempre aqui te espero.	1
@caserocosas: QUE NINGUN ORGANISMO DE DDHH.HAYA PLANTADO REPUDIO TOTAL POR LA MUJER DE CINCO ARGENTINOS.ES RARISIMO...	@caserocosas Si! MUY raro!	1
@caserocosas: QUE NINGUN ORGANISMO DE DDHH.HAYA PLANTADO REPUDIO TOTAL POR LA MUJER DE CINCO ARGENTINOS.ES RARISIMO...	@caserocosas Estos organismos perdieron total credibilidad de su lucha al politizarse.	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si, pero nunca gano nada aunque quiera	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Yo si, te aplaudo por todo lo que hacer por nosotros, gracias	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si, ya participo	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si, ya teniendo mucha fe en ganar	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Yo estoy participando para el iPhone 8	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si espero ganarme aunque sea 1	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Ya hice todo lo que pediste	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si claro	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel vamos que se puede	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si ya hice todo lo que indicaste	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si! ya estoy participando!!!!	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Gracias kartie te amo mucho mucho te amo mucho mucho ojala me lo gane	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Yo si suerte a todos	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Si te amo idola	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Ya Y Yo Quiero Uno!!!	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel yo si estoy participando en el sorteo	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Me lo quiero ganar para no gastar la plata de mis papas	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Ya estoy participando y espero ganarme uno...!	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel si ya estoy participando y espero ganar! Creo que me servirá mucho el Iphone	1
@IamKatieAngel: Ya estás participando para GANA E uno de estos 5 IP HONES 8777 ¿Qué esperas? Correleeee! http...	@IamKatieAngel Me encantan desde siempre y lastimosamente a mi me robaron el celular...	1

Table A56: Examples of labeled data: Twitter (21)

@ReaccionesATP: Un aplauso para ese guardia de seguridad que no sabía quién era Rafa Nadal	@ReaccionesATP @soyLuper esto te va a gustar	1
@JeonAmethyst: No se por que no les queda claro que Army es un fandom muy poderoso, cada que nos hacen algo nos hacemos tendencia porque...	@JeonAmethyst Y por eso casi no nos hacen caso..	1
@lara_santaella: ¿Sabías que además de hacer fotografía de calle, hago sesiones de fotos tó guapas? ¡Contratadme! ¡O dadle al menos!	@lara_santaella @Corvus_corax81 Me encantan tus fotos	1
@valtonyc: ¡DIFUSIÓN! Acaba de salir la sentencia, ¡me han condenado a 3 años y 6 meses de cárcel por las canciones! ón	@valtonyc Denunciado a europa	1
@valtonyc: ¡DIFUSIÓN! Acaba de salir la sentencia, ¡me han condenado a 3 años y 6 meses de cárcel por las canciones! ón	@valtonyc Lo siento mucho. Toda mi solidaridad.	1
@Xuxipc: ÚLTIMA HORA: en protesta por los encarcelamientos, Pedro Sánchez mañana le llevará el café y el Marca a Rajoy con cara muerta y ser...	@Xuxipc Muy en su línea. Jajaja	1
@monicacorralles: Esto es lo que pasa cuando se vota para "defender espacios" en una tiranía. Terminan ocupados por oportunistas arrastrando...	@monicacorralles @laidygomezf Yo creo que calladita es mejor, cada vez que hablas en favor de la gente del régimen, la cagas mas.	1
@DhasiaWezka: Hasta nunca amor.	@DhasiaWezka No creo eso yo	1
@kevsho9: Estoy por dar un show en el Luna Park y no lo puedo creer es para un evento privado, por eso no avisé nada, pero es hermoso...	@kevsho9 ay mi vida, te vas a llevar una gran experiencia en uno de los escenarios más importantes de argentina	1
@kevsho9: Estoy por dar un show en el Luna Park y no lo puedo creer es para un evento privado, por eso no avisé nada, pero es hermoso...	@kevsho9 Por que privado? Quiénes van a verlo?	1
@cap0: La hija de Manolo Escobar dice que su padre estaría a favor del derecho a decidir. Los comentarios a la noticia te sorprenderán.	@cap0 Yo no me creo que Manola fuera querido romper España, pero yo sabemos de la "pata" es esta señorita cojea	1
@cap0: La hija de Manolo Escobar dice que su padre estaría a favor del derecho a decidir. Los comentarios a la noticia te sorprenderán.	@cap0 ¿Y ahora qué van a cantar los abanderados si su canción protesta era el repertorio de Manolo Escobar?	1
@cap0: La hija de Manolo Escobar dice que su padre estaría a favor del derecho a decidir. Los comentarios a la noticia te sorprenderán.	@cap0 @tonialba Es que hasta que la hija opine les jode, di que si, de tu padre puedes hablar lo que te de la gana, faltaría más	1
@JesusCintora: La Policía no aclara si identificó a los fascistas que acosaron en Atocha a políticos catalanes	@JesusCintora Solo identifica. Bueno es imperdonable esta es la que tiene que defender el estado de derecho	1
@cap0: La hija de Manolo Escobar dice que su padre estaría a favor del derecho a decidir. Los comentarios a la noticia te sorprenderán.	@cap0 @carlosbardem Gracias por poner los comentarios a la noticia. Sabía que había catetos en este país, pero no pensaba que tantos	1
@infobae: Inédito, la canción de Santiago Maldonado sobre @Pontifex es y la Iglesia católica Por Luis Gasulla...	@infobae @Pontifex Mira @sergiomaldoh !!! Qué linda canción !!!	1
@AbrahamMateo: HOY ROMPEMOS EN EL LUNA PARK.	@AbrahamMateo Y VAYA SI LA ROMPISTE, BOMBÓN!!!!	1
@AbrahamMateo: HOY ROMPEMOS EN EL LUNA PARK.	@AbrahamMateo Fue hermoso estar una vez más apoyándote en este gran estadio	1
@SoyBienMamon: Regálame un besito más...	@SoyBienMamon Solo quiero darte un beso más	1
@SoyBienMamon: Regálame un besito más...	@SoyBienMamon Solo quiero darte un beso más...!	1
@ruggero: Soy la persona más feliz del mundo!	@ruggero Yo Soy feliz por ti	1
@wonhoftminhyuk: ESTO ME PONE LOS PELOS DE PUNTA Y LAS GRIMAS ME SALEN SOLAS SIEMPRE COMO EXPRESA TODO CON LA CARA. TRAIGANME PAÑUELOS...	@wonhoftminhyuk Esa actuación siempre me pone los pelos de punta es que es increíble	1
@DreambeachFest: de abono individual Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest Ojala haya suerte.	1
@DreambeachFest: de abono individual Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest Un poquito de suerte	1
@kevsho9: Estoy por dar un show en el Luna Park y no lo puedo creer es para un evento privado, por eso no avisé nada, pero es hermoso...	@kevsho9 En serio? Pensé que hoy te iba a conocer	1
@mangelrogel: @Rubiús Eres la única persona que conozco que se pone una canción personal de despertador.	@mangelrogel @Rubiús Dejalo, es especial.	1
@mangelrogel: @Rubiús Eres la única persona que conozco que se pone una canción personal de despertador.	@mangelrogel @Rubiús Yo lo hago, no me harta para nada.	1
@Bibliaenlinea: DIOS nunca se equivoca, sus planes son perfectos.	@Bibliaenlinea Es tan sabio, y la humanidad muchas veces faltos de fe... No te preguntes, porque a mi, Dios.	1
@Bibliaenlinea: DIOS nunca se equivoca, sus planes son perfectos.	@Bibliaenlinea Es algo muy especial	1
@MatiildaxD: Qué ser humano basura el que pide empanadas de verdura y después come de las buenas. Es el mismo que se hace el chavista de...	@MatiildaxD Se llenan de botox y se rellenan la boca, Hay de todo.	1
@DreambeachFest: de abono individual Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest A ver si me toca algo	1
@DreambeachFest: de abono individual Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest me metio en twitter antes de hacer los informes	1
@DreambeachFest: de abono individual Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest Ahora no me puedo salir hasta que confirmeis	1
@PaulaVazquezTV: Y mientras sacudes tu bandera ...	@PaulaVazquezTV Para los fanáticos de medidas sociales ¡¡¡viva el generalísimo!!!	1
@gabrielrufian: No tenéis vergüenza.	@gabrielrufian Pero realmente crees o piensas que algún día han tenido un atisbo tan siquiera de vergüenza?	1
@gabrielrufian: No tenéis vergüenza.	@gabrielrufian No nos callarás!!! CATALUNA ES ESPAÑA!!!	1
@gabrielrufian: No tenéis vergüenza.	@gabrielrufian Tu si que no tienes vergüenza, ojalá os bajen el sueldo	1
@villalobossebas: no corría de ustedes, corría con ustedes	@villalobossebas Siempre juntos	1
@villalobossebas: no corría de ustedes, corría con ustedes	@villalobossebas Yo soñé que me invitaste al cine y al ir caminando nos agarramos de la mano.	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Cristo siempre está en mí. Gracias a Dios	1

Table A57: Examples of labeled data: Twitter (22)

@gabrielrufian: No tenéis vergüenza.	@gabrielrufian Tu si que no tienes vergüenza	1
@MisOdios: El Papa no viene porque perdió Scioli en 2015 y porque perdió Cristina en 2017. Cualquier otra excusa es humo.	@MisOdios Y va a perder en el 2019	1
@MIAREsproject: Dulces sueños	@MIAREsproject solo si tu estas en ellos	1
@escorpio_hn: Para conquistar a se necesita paciencia, valor, cerebro, encanto y ser sencillo.	@escorpio_hn Todo eso si pero sobre todo sencillo	1
@Elaguijon_: Los mismos que hoy piden libertad para los golpistas e xigirán mañana paredón para aquellos que no comulgan con sus ideas.	@Elaguijon_ Tienes toda la razón pero tengo que decir que han des pertado muchos de los que tenían miedo y eso es fantástico	1
@Soymarioruiiz: Buenos díaaaaas	@Soymarioruiiz Buenos dias	1
@Soymarioruiiz: Buenos díaaaaas	@Soymarioruiiz Buenos dias bebé	1
@Soymarioruiiz: Buenos díaaaaas	@Soymarioruiiz Buenos	1
@Olaff_: Qué hermoso el verano, sentir cómo te transpiran los huevos, el cuello, el oje y los pies. Viajar hacia el laburo, en bondi....	@Olaff_ Quizas si tuvieras un BMW viajarías con aire y disfrutarías del verano.....	1
@patricialopezl: Ahora????	@patricialopezl @publico_es Si, claro, cuando Rajoy ya se ha pringa do hasta las cejas. No iba él a tener ninguna iniciativa!! útil	1
@patricialopezl: Ahora????	@patricialopezl Ya le están llegando los recados que se a vuelto a e quivocar !	1
@patricialopezl: Ahora????	@patricialopezl Estos han perdido el Norte y mientras lo pierden est án arrastrando a todos con ellos. Ni socialista...	1
@caroflores23: 3 películas en 1 día	@caroflores23 @PauPrinsesita tenemos que hacer esto	1
@caroflores23: 3 películas en 1 día	@caroflores23 Lo que queríamos hacer @Adriandlc83	1
@caroflores23: 3 películas en 1 día	@caroflores23 @braulio_barron lo hubieramos hecho	1
@caroflores23: 3 películas en 1 día	@caroflores23 @AndreaMolibux no pues a tus novios se les da muy bien	1
@caroflores23: 3 películas en 1 día	@caroflores23 @LizbethSalazarT Más logística para la otra	1
@caroflores23: 3 películas en 1 día	@caroflores23 @FridaaA21 estas 3 necesitamos ir a ver	1
SEGUIRME A MI Y A @competicontrol DAR Y A LA FOTO ...	@TOBBALink @CompetiControl Participo por si acaso	1
SEGUIRME A MI Y A @competicontrol DAR Y A LA FOTO ...	@TOBBALink @CompetiControl QUE EMOCIÓN!!!	1
@escriba: Atacan a Estela de Carlotto y preparan una ofensiva judici al contra todos los organismos de derechos humanos como adelant ó Cla...	@escriba @Gatosylvestre Organismos de que ?no una palabra de lo s 5 muertos argentinos en EEUU x isis	1
@escriba: Atacan a Estela de Carlotto y preparan una ofensiva judici al contra todos los organismos de derechos humanos como adelant ó Cla...	@escriba @Gatosylvestre Muy bien, si solo se la pasan defendiendo delinquentes como milagro sala	1
@escriba: Atacan a Estela de Carlotto y preparan una ofensiva judici al contra todos los organismos de derechos humanos como adelant ó Cla...	@escriba @Gatosylvestre Cómo tiene que ser...como van a pedir que le devuelvan los nietos estamos en de mo cra cia...en democracia	1
@gabrielrufian: No tenéis vergüenza.	@gabrielrufian @miqueli Si tuvieses verguenza dimitirías	1
@Cristina_H_: No es que esté de moda denunciar el acoso, es que A HORA hemos generado (con esfuerzo) un contexto social en el que podemos...	@Cristina_H_ Es que Ahora a cualquier cosa le llamas acoso claro q ue detrás de esto hay mucho interés por las asociacion	1
@Santi_Vatt: Un bebé punga	@Santi_Vatt Eso ya se ha visto	1
@ForretsGump: Estará el FGE Maza preparando la petición de prisió n incondicional y sin fianza para M. Rajoy o aún no sabe quien es? http...	@ForretsGump @jorpi9 No tiene tiempo, lo dedicara a comprar som brenitos	1
@ForretsGump: Estará el FGE Maza preparando la petición de prisió n incondicional y sin fianza para M. Rajoy o aún no sabe quien es? http...	@ForretsGump lo está afinando	1
@PibesResentidos: Cuando estás acostumbrado a que te cancelen	@PibesResentidos @rojoeli8 @CamiluZerda cuando te invitan a salir	1
@DjMaRiiO_90: Primer rival del día pero venga ya jajajajaja	@DjMaRiiO_90 A que hora es el directo?	1
@progadgetYT: [& FOLLOW] He llegado a 1.5M de suscriptore s y para celebrarlo:SO EO INTERNACIONAL iPHONE X Participa:	@progadgetYT felicidades por llegar a 1.5 millones	1
@progadgetYT: [& FOLLOW] He llegado a 1.5M de suscriptore s y para celebrarlo:SO EO INTERNACIONAL iPHONE X Participa:	@progadgetYT Espero ganar	1
@progadgetYT: [& FOLLOW] He llegado a 1.5M de suscriptore s y para celebrarlo:SO EO INTERNACIONAL iPHONE X Participa:	@progadgetYT Participó!!!	1
@progadgetYT: [& FOLLOW] He llegado a 1.5M de suscriptore s y para celebrarlo:SO EO INTERNACIONAL iPHONE X Participa:	@progadgetYT Listo	1
@progadgetYT: [& FOLLOW] He llegado a 1.5M de suscriptore s y para celebrarlo:SO EO INTERNACIONAL iPHONE X Participa:	@progadgetYT Ya lo e echo por Instagram y por aquí! Participo! Oj alá toque!	1
@elcitzizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitzizen Yo tambien lo pienso puro blabla y no los agarran como a otros Dictadores	1
@elcitzizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitzizen Ps de eso no ha pasado y mientras más días más se diluy e más se olvida	1
@elcitzizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitzizen @tatalibre Y así pareciera que es.	1
@elcitzizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitzizen Hasta los momentos van teniendo la razón	1
@RafaMorenoValle: Necesitamos garantizar gobernabilidad y mayori as estables en el congreso, por ello la importancia de un ...	@RafaMorenoValle @Trelevision @NOTIRIVAS @YucatanALaMano @ReflejoPublico @YUCATANALMINUTO La democracia es simbolo...	1

Table A58: Examples of labeled data: Twitter (23)

@CapitanCaceres: ESTATUTO d ROMA dice Textualmente q es LESA HU MANIDAD los Ataque a POBLACIÓN CIVIL x parte d ORGANIZACIONES ERP-Montone...	@CapitanCaceres Totalmente de acuerdoBasta deQuedar bien (simpati co)Que hacer bien (antipatico) Justicia Pareja...	1
@CapitanCaceres: ESTATUTO d ROMA dice Textualmente q es LESA HU MANIDAD los Ataque a POBLACIÓN CIVIL x parte d ORGANIZACIONES ERP-Montone...	@CapitanCaceres @aagaetan Capitan ya los juzgaran: y cómo hará con los desaparecidos?	1
@CapitanCaceres: ESTATUTO d ROMA dice Textualmente q es LESA HU MANIDAD los Ataque a POBLACIÓN CIVIL x parte d ORGANIZACIONES ERP-Montone...	@CapitanCaceres Yo no lo puedo entender como se puede subsidiar a lo subversivos. Cuando mataron gente civiles.	1
@fckyongji_: BTS se dará cuenta con los años, así como lo hizo GD qu e esos premios no valen nada# #_BY BTS...	@fckyongji_ @Jaquikookie Esto ya ha pasado con Super Junior varias veces les quitaron premios para dárselo a otros...	1
@AldoDuqueSantos: Y continúa el triste periplo de inauguraciones.Com o si con 10 días de inauguraciones pudiera cambiar 4 años de inefic...	@AldoDuqueSantos Que les parece si nuestra presidenta se traslada en transantiago para las inauguraciones que haga...	1
@AldoDuqueSantos: Y continúa el triste periplo de inauguraciones.Com o si con 10 días de inauguraciones pudiera cambiar 4 años de inefic...	@AldoDuqueSantos @rfanaspicon Amermelado! Significa que no hubo ineficiencia!! Se inaugura lo que se ha trabado y bien trabajado.	1
@PatriotaNene: Argentinos que han venido a España y se permiten el l ujo de querer trocear nuestra Nación, si quieres la deportación d...	@PatriotaNene Que independicen la Patagonia.	1
@ElJuanpaZurita: Buen bello día	@ElJuanpaZurita Si un buen día para estar con tu Familia	1
@ElJuanpaZurita: Buen bello día	@ElJuanpaZurita Igualmente precioso!!!	1
@ElJuanpaZurita: Buen bello día	@ElJuanpaZurita Bello día para ti	1
@ElJuanpaZurita: Buen bello día	@ElJuanpaZurita Buen bello día!	1
@ElJuanpaZurita: Buen bello día	@ElJuanpaZurita Hola!	1
@tattopenagos: Buenos días !!! Como amanecen, hoy es un día especia l por qué es el día feliz día señoritas	@tattopenagos @mariobautista_ Feliz día para ti tambien	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Me encantan los amo	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Rtm por favor saludame me encanta n muchos tus videos	1
@dosogas: A quien le haremos la videollamada la próxima semana? (D e todas las redes sociales, los más activos)	@dosogas Yo quiero	1
@dosogas: A quien le haremos la videollamada la próxima semana? (D e todas las redes sociales, los más activos)	@dosogas Quiero verles	1
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani En serio??	1
@julenaretakoa: La policía concibe el himno de España como método de tortura a vascos y catalanes. Esclarecedor. ñaEsTorturahttps...	@julenaretakoa Es la única canción de amor que conocen, nos quieren seducir. Dime qué cantas y te diré quien eres	1
@julenaretakoa: La policía concibe el himno de España como método de tortura a vascos y catalanes. Esclarecedor. ñaEsTorturahttps...	@julenaretakoa @jordiborras Razón no les falta. Lo que les jode hasta no poder más	1
@EAristeguieta: Al colombiano Maduro y al sirio El Aissami no les pued e doler Venezuela. Lo vergonzoso es que a nuestros militares tampo...	@EAristeguieta Militares en Venezuela donde? Una partida de arrastrad os	1
@PastorCashLuna: Si declaras con tu boca que Jesús es el Señor y cree s de corazón que Dios lo levantó de entre los muertos, Dios te salv...	@PastorCashLuna Amén	1
@PastorCashLuna: Si declaras con tu boca que Jesús es el Señor y cree s de corazón que Dios lo levantó de entre los muertos, Dios te salv...	@PastorCashLuna Amén eternamente. Lo declaró gracias a Dios. Inscrit os en el libro de la vida del Cordero de Dios	1
@Cornudoyucatan: Hay video??? Si!!!!!!!	@Cornudoyucatan Por fin!!!! Que larga se hizo la espera putita precios a para volver a verte en acción. Grabó el cornudo?	1
@Heberizeth: Luego de su trabajo de investigación en Tocorón @jesus medinae recibió amenazas. Hoy desapareció en la tarde ...	@Heberizeth @SabrinaMartinR @jesusmedinae El mundo debe saber que estamos subyugados por un régimen TERRORISTA, al...	1
@Argengchogau: AYER SABADO por la tarde Río Gallegos,CFK saliendo por puerta de servicio, DESCUBIE A, vecinos le gritan chorra, ladro...	@Argengchogau "Los" vecinos,o "LA" vecina?	1
@GirealaDerecha: Este NarcoCandidato evadió durante 20 años sus obl igaciones tributarias y hoy ofrece subir impuestos para ayudarlo a Ud...	@GirealaDerecha @euge2145 pero por supuesto si quitar la ajena sin n ingún tipo de remordimiento es muy facil	1
@alinavarro88: Paco Alcácer sólo lleva un gol menos (10) que Karim Be nzema (11), en lo que va de 2017 en 1733 minutos menos.	@alinavarro88 @Bajarlaalpasto Según Zizou Benzema es el mejor 9 del mundo.	1
@alinavarro88: Paco Alcácer sólo lleva un gol menos (10) que Karim Be nzema (11), en lo que va de 2017 en 1733 minutos menos.	@alinavarro88 @Jorgerdz10Jorge Espero que Valverde rectifique su err or en descartar a Paquito y le de más minutos	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @Lupita67897473 @KimberlyLoaiza_ Me encanto el vid eooo Espero y un día me saluden en su video, sería que algo tan herm oso	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Se Ven bien Juntos	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ siempre saben cómo sacarnos una s onrisa	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ ! Los Amo.Los Dios Siempre Me Saca n Una Sonrisa. Los Adoros!	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Que lindos se ven	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ @KimberlyLoaiza_ @Juandedios_P P erfectos. Muchas bendiciones.	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Me encantan..	1
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracia s @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Son los mejores del mundo mundial! Los amo muchísimo!	1
@ZOUISEOK: Harta estoy de que no reciban lo que merecen, el grupo tiene un potencial increíble para que los dejen tirados... TreatG...	@ZOUISEOK No entiendo, me explicas?	1
@OscarOrtizG: 1. Solicitamos a @CCConstitucional ratificar requisitos par a participación política de farc, ordenados en sentencia C577 20...	@OscarOrtizG @CCConstitucional Excelentes conclusiones, deben hacers e publicas	1
@OscarOrtizG: 1. Solicitamos a @CCConstitucional ratificar requisitos par a participación política de farc, ordenados en sentencia C577 20...	@OscarOrtizG @CCConstitucional He aquí un poco de lo que interesa e n cuanto al tema de la participación política de FARC @AlfaroAguas	1

Table A59: Examples of labeled data: Twitter (24)

@BeatrizTalegon: La población Catalana tiene una cultura democrática m	@BeatrizTalegon @Jadouken10 Por desgracia, la reacción del resto de Es	1
uy por encima de la media española. Por eso aquí no se entiende ca...	pana en estas semanas le da la razón	
@BeatrizTalegon: La población Catalana tiene una cultura democrática m	@BeatrizTalegon Que manía con generalizar, que unos borregos indepes	1
uy por encima de la media española. Por eso aquí no se entiende ca...	no entiendan que saltarse la Ley tiene sus consecuen...	
@BeatrizTalegon: La población Catalana tiene una cultura democrática m	@BeatrizTalegon Deja de insultar a la ciudadanía. Quien est demostrando	1
uy por encima de la media española. Por eso aquí no se entiende ca...	que no tiene cultura democrática eres tú...	
@BeatrizTalegon: La población Catalana tiene una cultura democrática m	@BeatrizTalegon Gracias Bea. En concreto, la media Catalunya contra la i	1
uy por encima de la media española. Por eso aquí no se entiende ca...	ndependencia hace elevar la nota.	
@BeatrizTalegon: La población Catalana tiene una cultura democrática m	@BeatrizTalegon En serio? Pues gracias oye, porque sabes? Soy catalán, y	1
uy por encima de la media española. Por eso aquí no se entiende ca...	una cosa te digo, lo entiendo todo.	
@BeatrizTalegon: La población Catalana tiene una cultura democrática m	@BeatrizTalegon @roccasagran Tienes toda la razón...y vas a recibir fuerte	1
uy por encima de la media española. Por eso aquí no se entiende ca...	por decirlo. Animo y gracias	
@soycrismorena: Gracias @elotroladook, la risa sanadora provocada por	@soycrismorena @ElOtroLadoOk Que semillero de talento. Y no solo eso.	1
excelentes y quendos actores en una comedia que no se pueden perd...		
@soycrismorena: Gracias @elotroladook, la risa sanadora provocada por	@soycrismorena @ElOtroLadoOk Que grupo humano. Debe sentirse un a	1
excelentes y quendos actores en una comedia que no se pueden perd...	mor intenso	
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Y agradecido siempre fiel a tu lado va a estar!!!! Ese am	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...	or no tiene precio!!!	
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Yo también rescate mi perrita de la calle. Ya casi 3 años	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...	Más bella!!!	
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Son ellos los que cuando te sentis solo te acompañan s	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...	iempre...el te lo va a agradecer siempre !!!	
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Que gran persona eres! Que Dios te bendiga	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...		
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Que bueno lo tuyo yo tengo tres y todos rescatados de	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...	la calle son un amor	
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Que lindoooo !! Los callejitos rescatados son los mas	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...	amorosos !	
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Que hermoso gesto! Felicitaciones! alegraste el día!	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...		
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 @gastonmarraOK Son los más agradecidos	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...		
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 @gastonmarraOK Aplausos muchos aplausos	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...		
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 q suerte tienen ambos, disfrútense	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...		
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 que carita dulce!!!! Que alma buena debes ser.	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...		
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 No hay nada más fiel	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...		
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer	@missencuestas2 Felicitaciones! Yo también adopté uno de la calle y es u	1
to en un tacho de basura.hoy es mi compañero fiel donde yo estoy...	n perro muy feliz.	
@FSerranoCastro: En este país es más peligroso rebelarse contra la dictad	@FSerranoCastro Del género vive demasiada gente ya. Crear negocio de	1
ura de género que alzarse en sedición y rebelión contra el Es...	generar una masa de procesados tratando igual...	
@libertaddigital: Documentos revelan que el plan de Generalidad pasaba	@libertaddigital @conlluvia Hacerlo público!!!!	1
por controlar aeropuertos, puertos, centrales nucleares y... https...		
@MashiRafael: ¡Qué triste momento que vive el país! Anti valores por do	@MashiRafael Así es mi presidente pero aquí tienes a los leales ante tant	1
quier.	a mediocridad y deslealtad un abrazo y H.L.V.S comandante Fafael	
@imyashua: Quisiera tenerte y darte todoY verte crecer como una floral l	@imyashua Ayy Me encantas	1
ado mio darte gotas de amor.		
@imyashua: Quisiera tenerte y darte todoY verte crecer como una floral l	@imyashua Quisiera que vengas a peru para verte y abrazarte fuertísimo	1
ado mio darte gotas de amor.	y decirte lo mucho que te quiero	
@imyashua: Quisiera tenerte y darte todoY verte crecer como una floral l	@imyashua AWWW TU ERES MUY AMOROSO	1
ado mio darte gotas de amor.		
@imyashua: Quisiera tenerte y darte todoY verte crecer como una floral l	@imyashua Yo quisiera conocerte ,abrazarte y nunca soltarte	1
ado mio darte gotas de amor.		
@imyashua: Quisiera tenerte y darte todoY verte crecer como una floral l	@imyashua Se que eso no va para mi pero que lindo	1
ado mio darte gotas de amor.		
@PinkyGrace_ Yo aquí veo cadena perpetua. No sé vosotros.	@PinkyGrace_ @Im_ruben Cadena perpetua me parece muy poco, mínim	1
	o durante toda la vida comer tres veces al día esa...	
@WikiChelsea: Video: El gol de Alvaro Morata gracias a la asistencia de C	@WikiChelsea @denisvox Increible!!!	1
esar Azpilicueta.		
@omnium: Madrid, muchas gracias por salir a la calle para pedir libertad	@omnium Gracias Madrid,la falta de democracia afecta a todos los habit	1
y democracia itics	antes de esta Península!	
@omnium: Madrid, muchas gracias por salir a la calle para pedir libertad	@omnium Madrid sera la tumba del fascismo.	1
y democracia itics		
@Soymarioruiz: Buenos diaaas	@Soymarioruiz Buenos días	1
@alejandrosanmo: Este chito ha pedido por favor que no se difundan ni	@alejandrosanmo @1_jam2 No pasa nada. Tienen cancha para hacer lo q	1
su nombre ni sus fotos. Se llama Antonio Manuel Guerrero Escudero....	ue quieran. Ya lo vimos el 1 octubre. Igual que...	
@bossyviciconte: "Gracias por soñar, por creer, por luchar, por ser siemp	@bossyviciconte @MicaViciante Lo vio, felicitaciones! Te lo mereces	1
e vos misma @MicaViciante"		
@pburelli: A @FreddyGuevaraC le cobran el liderazgo valiente que ejerció	@pburelli @FreddyGuevaraC Es valiente enviar jóvenes a la muerte para l	1
durante 130 días de calle que cambiaron 180° posición del mundo....	uego negociar con el gobierno.	
@ElNinodeltabaco: Cuando la ensalada no para de contarte chistes	@ElNinodeltabaco Por favor si puedes me mandas los ingredientes	1
@alphabettfredy @freddyleyva Me faltó mencionarlo por aquí xdx @fre	@freddyleyva en serio debes mirar esto esta increíble miralo!	1
ddyleyva ¡MIRA ESTOOO!		
@alphabettfredy @freddyleyva Me faltó mencionarlo por aquí xdx @fre	@alphabettfredy @freddyleyva de verdad deseo que veas esto @freddyle	1
ddyleyva ¡MIRA ESTOOO!	yva :(

Table A60: Examples of labeled data: Twitter (25)

@DanielaRued: No puedo con Selena y Justin, me va a dar un infarto	@DanielaRued Son tan lindo verdad	1
@HambreCero: A veces,ser honesto,es ser un revolucionario...	@HambreCero A veces no. SIEMPRE	1
@anabelenroy_tve: ¿Cuánto tenemos que aprender de ellos?... Qué delicia de video	@anabelenroy_tve @sevilajordi Maravilloso, pero nada sorprendente	1
@anabelenroy_tve: ¿Cuánto tenemos que aprender de ellos?... Qué delicia de video	@anabelenroy_tve Mi vida!! No hay cosa más tierna que ver estás imágenes	1
@anabelenroy_tve: ¿Cuánto tenemos que aprender de ellos?... Qué delicia de video	@anabelenroy_tve Muy buen video y lleno de humanidad.	1
@cacoinfante: Terminé de pintar la casa, apareció Bachelet, y la inauguró	@cacoinfante qué suerte!! hubiese sido el ñiña @sebastianpinera te la remodela y no le pone puertas ni ventanas	1
@kazan2418: Ridículo de la prensa española preguntando al portavoz de la fiscalia, si los politicos catalanes iban esposados.	@kazan2418 Porque ridículo es un criminal no?	1
@fedevigevani: Digo esto porque no me gustan los domingos	@fedevigevani A mi me gustas tú	1
@BeatrizTalegon: La población Catalana tiene una cultura democrática muy por encima de la media española. Por eso aquí no se entiende ca...	@BeatrizTalegon Tienes razón, y punto	1
@BeatrizTalegon: La población Catalana tiene una cultura democrática muy por encima de la media española. Por eso aquí no se entiende ca...	@BeatrizTalegon Tu es que eres muy intelectual.	1
@DavidCardona_5: Cuando alguien de Facebook llega por primera vez a Twitter:	@DavidCardona_5 Soy yo jajaja!	1
@JOSEMANUELSTO1: Franquistas en España no quedan, apestan mucho mas los chavistas, castistas, leninistas q hay cobrando sueldazos p...	@JOSEMANUELSTO1 @elrancius Quizas los chavistas no se duchen,tan cierto como que habeis muchísimos fascistas en España	1
@JOSEMANUELSTO1: Franquistas en España no quedan, apestan mucho mas los chavistas, castistas, leninistas q hay cobrando sueldazos p...	@JOSEMANUELSTO1 Además de verdad	1
@_infoLibre: Pons acusa al ministro de Interior belga de pertenecer a un partido "xenófobo" tras cuestionar la justicia española. https://un...	@_infoLibre Increible !!!	1
@soffiaaa: Que lo más importante del avance tecnológico de un teléfono sea emojis cantando, resume a la raza humana hoy en día. ...	@soffiaaa Me encanta como todos los q se supone saben de tecnología te cagan! Si no entienden que en principio tu humor no hay nada que hacer!	1
@atajoslargos: En el hotel, pedi pasta de dientes. No sé si usarla o cuidarla y verla crecer y apoyarla en las decisiones que tome en su...	@atajoslargos Usala. Después se seca.	1
@okdiario: Reúnen 44.000 firmas para pedir que Echenique sea declarado "persona non grata" en España	@okdiario Dónde hay que firmar?	1
@daloest10: "No hay ningún video ni foto que pruebe que el niño fue penetrado por mi cliente, así q es inocente" (Eduardo Franco Looor) PE...	@daloest10 Increible, de verdad dijo eso ese abogado del Glas? Que será que tienen que en lugar de sangre esa gente.	1
@Camila_Cabello: Te amo.....	@Camila_Cabello VI EL VIDEO Y ME EMOCIONÉ TANTO CANTAS CON TODA EL ALMA Y ESO ME ENCANTA TE AMO MI BEBE	1
@Coordinadora253: Madrid está con el pueblo catalán. Desde la Puerta del Sol ósAlRegimenDel78	@Coordinadora253 Con el pueblo catalán NO, con la parte indepe del pueblo catalán, que es diferente.	1
@MargaritaRepo: Sin comentarios...	@MargaritaRepo @toro_abogado Se necesita ser muy ignorante para decir tantas babosadas.	1
@auronplay: Por fin me han puesto los 280 caracteres soy muy feliz en este momento quiero mandar un saludo a mi abuela a ver si un día v...	@auronplay Tu regalo de cumple por parte de Twitter Por cierto, felicidades crack!	1
@auronplay: Por fin me han puesto los 280 caracteres soy muy feliz en este momento quiero mandar un saludo a mi abuela a ver si un día v...	@auronplay Pues los de Twitter quisieron hacerte un regalo	1
@Profeta_Baruc: El hilo de hoy comienza con una pregunta: ¿Qué tienen en común estas tres fotos?	@Profeta_Baruc Me encantan estos hilos! Muchas gracias!	1
@yoonmincake: Mirar como Yoongi se rie al acercarse a Jimin y luego cuando se va Jimin se rie, my heart	@yoonmincake @LuzAoSora Me encanta como se arregla el cabello jimin una vez que Yoongi se va.	1
@Camila_Cabello: Te amo.....	@Camila_Cabello Me encantas.	1
@bryanmouque: Todos queremos libertad.	@bryanmouque A veces no te entiendo	1
@mohagerehou: 60.000 personas pidieron en Polonia que Europa sea solo blanca. Por esto tenemos que llenar las calles contra el racismo....	@mohagerehou Esto es muy preocupante...	1
@Camila_Cabello: Te amo.....	@Camila_Cabello Te amo MUCHO MÁS	1
@Camila_Cabello: Te amo.....	@Camila_Cabello Yo también	1
@Camila_Cabello: Te amo.....	@Camila_Cabello te amo mucho	1
@Orion_fighter: Belkys Solorzano ayer deja testimonio de la que vive n...	@Orion_fighter Que tristeza. Esto no puede pasar!! No lo acepto!	1
@Camila_Cabello: Te amo.....	@Camila_Cabello Quiero conocerte	1
@Camila_Cabello: Te amo.....	@Camila_Cabello Y yo a ti	1
@Camila_Cabello: Te amo.....	@Camila_Cabello Te amo más	1
@Camila_Cabello: Te amo.....	@Camila_Cabello yo también TE AMO	1
@Nega_Maiz: Los chavales de Altsasu cumplen hoy un año en la cárcel, 365 días con sus noches, prisión preventiva a la espera de juicio p...	@Nega_Maiz Feliz aniversario!! Que cumplan muchos más!!	1
@sugaobsession donde lo puedo ver ??	@sugaobsession Los siete son tan tímidos eso me encanta	1
@sugaobsession donde lo puedo ver ??	@sugaobsession Los siete son tan tímidos eso me encanta	1
@KochMedia: es Su velocidad y su estilo de optimismo	@KochMedia es Me encantaba "Sonic the hedgehog " de la Mega Drive.Lo mejor? La música y cmo caía rodando con ese ru...	1

Table A61: Examples of labeled data: Twitter (26)

@mbachelet: Es que fui como mamá, no como Presidenta, Franci	@mbachelet Me encantaría tener una foto con mi Presidenta	1
sco. ¡Bonita selfie! Cariños.		
@yuzucchis Esto me recuerda a cierto profesor mío que llega 10 minutos tarde en cada examen y luego se queja de que...	@yuzucchis @Vocaloid365 En serio? Que el año que viene voy!	1
@FlorHalfon: Nadie le hizo a Leuco un comentario antisemita. Le gritaron: "¡Hablá de DyN y de Magnetto, forro!". Estuve ahí. Vi la situa...	@FlorHalfon @tunkollmann Muy bueno	1
@Pantomima_Full: Hemos hecho daño a un chaval con nuestro video PROYECTOS. Rayados y pensando en retirarlo de las redes. No queremos que...	@Pantomima_Full No os paséis, el muchacho necesita un abrazo. Solo pide cariño	1
@TomiOlava: Yo no quiero ser un opositor irracional pero realmente no puede un presidente sacar un lateral de esta manera...	@TomiOlava @FeeedeRP Un hombre !!! no puede sacar un lateral así. No tiene barrio calle Nada	1
@elcosodelapizza: MATEMÁTICA QUE MATERIA DEL O O LA PUTA QUE TE RE MIL PARIÓ Y VOS TAMBIÉN FÍSICO QUÍMICA LA CONCHA DE TU MADRE NO ENTI...	@elcosodelapizza Parece que alguien se la llevo	1
@_fede_ok: Me dan un asco las minitas así. Todas sus fotos mostrando el tremendo lomo q tiene pero dice que tiene un "cuerpo de mierda"...	@_fede_ok Por fin alguien lo dijo	1
@_fede_ok: Me dan un asco las minitas así. Todas sus fotos mostrando el tremendo lomo q tiene pero dice que tiene un "cuerpo de mierda"...	@_fede_ok @Maaalegonzalez Alguien lo tenía que decir !	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Hola bebe buenos días!!!	1
@pardodevera: Después de conocer esta grabación, a los votantes del PP les quedan tres opciones:- admitir que prefieren a un partido c...	@pardodevera Después de esto todavía dice @marianorajoy que está fuerte. votantes del PP en vuestras manos está.	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Muy bien día te amo	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buen día bebé	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buen día	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Lindo día	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buen día como estas que haces?	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Cómo estás???	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel BUEN DÍA HERMOSO	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Te amo, buenos días bebé.	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buenos días ¿como amaneciste? ahre te amo	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel buen día mi amor	1
@nathcampost: Buenos días! Y feliz día a todas las Bautisters	@nathcampost Feliz día a ti también	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Cómo estás?	1
@pccomponentes: VAMOS QUE SOLO TIENES HASTA MAÑANA PARA PARTICIPAR! Comenta con y sigue a PcComponentes	@pccomponentes @MSI_IBERIA a cruzar los dedos!	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buenos días bello	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buenos días guapo	1
@CristinaSegui: El próximo sábado 18 de noviembre hay que llenar. Sol para reivindicar ónYA Justicia salarial para nuestra po...	@CristinaSegui @Jusapol Muchas gracias por tu ayuda!	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buenos días hermoso Anoche fue increíble, en serio, muchas gracias	1
@ratioessendi: Piñera absolutamente contra las cuerdas y descontrolado cuando Mónica Rincón y Matamala le enrostran la evasión de impuestos...	@ratioessendi @DedierAlonso Lo increíble que es inmutable , nada le afecta , sigue su discursillo !	1
@AlbanoDante76: Si alguien descubre esta maravilla absoluta gracias a este tuit, habrá valido la pena abrirse una cuenta de Twitter...)	@AlbanoDante76 Me encanta!!!La tengo en Spotify!!	1
@_anapastor_: No me canso de verlo...	@_anapastor_ Por eso lo dijo	1
@_anapastor_: No me canso de verlo...	@_anapastor_ Ya me gustaría encontrármelo cara a cara, para zamparle un escupitajo entre las cejas.	1
@Anitta: Hola queridos! Envíenme sus preguntas! Te imaginas si respondo la tuya? Participa!	@Anitta Tienes muchos fans en Paraguay.. Ojalá aparezcas un día por aquí !!	1
@Anitta: Hola queridos! Envíenme sus preguntas! Te imaginas si respondo la tuya? Participa!	@Anitta @danditeixeira Tienes pensado venir a España o hacer un disco en Español??? te amo	1
@HassNassar: Si los magistrados de la JEP y los miembros de comisión de la verdad tienen la misma narrativa del conflicto, no habrá just...	@HassNassar @MariaFdaCabal Cual es el temor a la JEP y a la Comisión de la verdad?	1
@HassNassar: Si los magistrados de la JEP y los miembros de comisión de la verdad tienen la misma narrativa del conflicto, no habrá just...	@HassNassar @carlitocorrearr Cómo puedes aseverar eso?	1
@bryanmouque: Me encantaría poder dormir.	@bryanmouque Me encantaría que me hicieras caso.	1
@bryanmouque: Me encantaría poder dormir.	@bryanmouque Me encantaría poder abrazarte	1
@broxoxmiswebs: Dice el refrán: cuando el río suena, es que agua lleva, así como los rumores de la nómina de los Zetas.¡Sorpresas te da...	@broxoxmiswebs Y saben que va a pasar??Nada!No hacemos nada, es mas facil darse la vuelta y decir que nada pasa!	1
@broxoxmiswebs: Dice el refrán: cuando el río suena, es que agua lleva, así como los rumores de la nómina de los Zetas.¡Sorpresas te da...	@broxoxmiswebs Y que se hace al respecto? Bola de corruptos y tapaderas en todos lados por eso estamos como estamos	1
@soylannavarro: Algunos de los mejores momentos de la vida son los que no puedes contarle a nadie.	@soylannavarro Como conocer a ustedes, como entre al fando m. como me perdí en sus éxitos.	1

Table A62: Examples of labeled data: Twitter (27)

@JuanOrlandoH: El trabajo sin descanso, siempre rinde resultados; Ma ry Portillo nos lo demuestra con excelente participación en Campeona te	@JuanOrlandoH Felicidades eso es gracias a tu enorme trabajo adelan te	1
@JuanOrlandoH: El trabajo sin descanso, siempre rinde resultados; Ma ry Portillo nos lo demuestra con excelente participación en Campeona ...	@JuanOrlandoH Muchas Felicidades, su excelente trabajo ha dado bue nos resultados destacando ...	1
@JuanOrlandoH: El trabajo sin descanso, siempre rinde resultados; Ma ry Portillo nos lo demuestra con excelente participación en Campeona ...	@JuanOrlandoH Muchas felicidades gracias por poner en alto el nomb re de nuestro país el trabajo sin descanso ...	1
@JuanOrlandoH: El trabajo sin descanso, siempre rinde resultados; Ma ry Portillo nos lo demuestra con excelente participación en Campeona ...	@JuanOrlandoH Dios bendiga mi gente gracias presidente sin su apoy o este logro no sería posible siga con su apoyo a los sectores producti vos	1
@CNCOMusic: ... Solo 3 días para los !	@CNCOMusic Ese premio es suyo bebés	1
@CNCOMusic: ... Solo 3 días para los !	@CNCOMusic Vayan listos que seguro seguro suben a recibir ese pre mio	1
@CNCOMusic: ... Solo 3 días para los !	@CNCOMusic Van a ganar !!! Ustedes se merecen todo	1
@CNCOMusic: ... Solo 3 días para los !	@CNCOMusic Vana ganar, se lo merecen!!!!!! los quiero mucho!!	1
@CNCOMusic: ... Solo 3 días para los !	@CNCOMusic Solo a tres días mis amores mucha suerte	1
@codevida: Belkis Solorzano perdió su trasplante (riñón) por falta de medicamentos. Falleció el mismo día que dio su testimonio. ...	@codevida @venezolanoenpie Que Tristeza e impotencia se siente	1
@natijota: Qué cagón el que te rompe el corazón de a poquito.	@natijota Vos sos como las hormigas rubia..no tenes ni sangre ni cora zón.!	1
@CNCOMusic: ... Solo 3 días para los !	@CNCOMusic ES SUYO LO SE	1
@JoseColinaP: En Miami hay que dejar la sinvergüenza de estar apo yando a artistas que son chavistas o conviven con el chavismo, por ej ...	@JoseColinaP Exacto! Vergüenza al que vaya!	1
@MiileBiebs: VISTE LAS MINAS QUE SE HACEN COMO UN RODETE E N LA CABEZA PERO NO ES UN RODETE BIEN HECHO ES UN RODETE HECHO PARA LA MIERDA...	@MiileBiebs Yo intento hacer esas mierdas y parezco mi maestra wth	1
@MarcosWitt: "Ore al Señor, y él me respondió: me libró de todos mis temores. Los que buscan su ayuda estarán radiantes de alegría: ning...	@MarcosWitt Amén. Así es	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @CarlosEsc_42 Que tío mas sinvergüenza !!!!!	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @JuanmaRubio8 No le defiendo... pero lo que dice no es lo que pones.	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @Albert_Rivera Su único error es confundir Griso por Gris ó. Manipulación de pacotilla y tonto quien te crea.	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @Albert_Rivera Esta jente son basura no merecen la pena y los que le sigen pues que tiren a ver donde llegan	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @Granpotrenco O sea que quieren la indepencia porque espanya les obliga a limpiar y no les de ha sentarse en los autobuses	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @Albert_Rivera Que fuerte!!! están locos!!!!!!	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @Albert_Rivera Y estos son los que quieren la mierda per dón la indep que cinicos son baturros y cansinos...	1
@Schuma78: ¡¡BRUTAL!!: Susanna Griso rebate los argumentos a lider independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @Albert_Rivera No manipules las dice : si ud. Piensa de es ta manera ud. Estaría en casa limpiando y hacien...	1
@RapMonsterMx: Recuerden que hoy @BTS_twt se presentará en The Last Last Show 5:3 0pm. (Hora México)Les dejo el link...	@RapMonsterMx @BTS_twt Ya me confundí es a la 1:30 o a las 6:30?	1
@Alberto 51: En la oposición venezolana, TODOS (tanto líderes como seguidores) creen tener la verdad absoluta en sus manos.Mientras pi...	@Alberto 51 Tienes toda la razón y lo peor es que crea inamovilidad mientras tanto nos va arrebatando los espacios posibles	1
@Rubiu5: Doblaje Español confirmed	@Rubiu5 No entiendo nada	1
@Rubiu5: Doblaje Español confirmed	@Rubiu5 Jajajaja me estoy muriendo de risa el mejor puto doblaje co n ese me quedo yo	1
@Rubiu5: Doblaje Español confirmed	@Rubiu5 Buenísimo	1
@Pedro_Arancon: Es una pena que los cientos de banderas desplega das en las calles estos días al grito de no hayan recibido ho...	@Pedro_Arancon Deprimente. Ni una sola bandera de España una gra nde y libre.	1
@Pedro_Arancon: Es una pena que los cientos de banderas desplega das en las calles estos días al grito de no hayan recibido ho...	@Pedro_Arancon @varelavito Muy buena reflexión.	1
@pccomponentes: SOMOS 3º A NIVEL NACIONAL EN TENDENCIAS! C omenta con y sigue a PcComponentes y @OMENbyHP y mu...	@pccomponentes @OMENbyHP necesito un regalo por mi cumpleaños s por parte de los mejores	1
@pccomponentes: SOMOS 3º A NIVEL NACIONAL EN TENDENCIAS! C omenta con y sigue a PcComponentes y @OMENbyHP y mu...	@pccomponentes @OMENbyHP Dios a ver si me toca	1
@pccomponentes: SOMOS 3º A NIVEL NACIONAL EN TENDENCIAS! C omenta con y sigue a PcComponentes y @OMENbyHP y mu...	@pccomponentes @OMENbyHP a ver si hay suerte	1
@Rubiu5: Doblaje Español confirmed	@Rubiu5 que buen doblaje donde te contrato para que doubles el ani me de virtual hero? ahre	1
@Rubiu5: Doblaje Español confirmed	@Rubiu5 al final vas a hacer vos el doblaje de tu anime??????	1
@PabloMM: 5 tiparracos que violaron a una cría de 18 años en un por tal piden respeto a su intimidad.	@PabloMM @XSalaImartin Me apunto para violarlos a ellos y que me respeten la intimidad	1
@Virginni: Si te violan 5 salvajes pero uno de ellos es GC el juez amit e como prueba un informe de un detective que espía a la víctima...	@Virginni @SusodeToro1 Ahora es cuando se echa de menos el fam oso "a por ellos". Pero claro, ahora no toca.verdad?	1
@8choOficial: Termina esta frase:Hoy hay _____	@8choOficial Pizza para cenar	1
@iFresaFria: Rt y vas a aprobar todo lo de este mes	@iFresaFria @JJAbuson @JJAbuson no me creo que lo hagas por que te lo creas JAJAJA	1
@el_pais: El juez, en una sentencia contra un maltratador de animales: "El que maltrata a un animal también despliega su instinto agresi...	@el_pais Claro, porque a los animales sólo los maltratan los hombres.	1

Table A63: Examples of labeled data: Twitter (28)

@LSisco17: Sólo espero que la RAE acepte "Chavista" como sinónimo de mamagüevo.	@LSisco17 Vergación, si todo mamawebó es chavista, con seguridad no hay marico que se libere de tal málcula.	1
@LSisco17: Sólo espero que la RAE acepte "Chavista" como sinónimo de mamagüevo.	@LSisco17 eso espero amiga jaja	1
@labeasanchez: Y pensar que todavía hay gente que cree que la única forma de comprometerse con algo es pagando...	@labeasanchez @fersuperlativo Y pensar que hay gente que votará por @labeasanchez	1
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana... :)	@ecosdelbalon @XabiAlonso Con el Carnicero de Tolosa? Y así, sin seguridad privada ni nada, héroes!	1
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana... :)	@ecosdelbalon @SpheraSports @XabiAlonso Madre mía uno de mis canales y jugadores predilectos.	1
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana... :)	@ecosdelbalon @XabiAlonso Sin ninguna duda, que envidia y a la vez que alegría ver como vais creciendo chicos	1
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana... :)	@ecosdelbalon @XabiAlonso Wooooaaahhh. Que envidia más sana. Y que emoción verlos donde están ahora. Saludos!	1
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana... :)	@ecosdelbalon @XabiAlonso Estaremos atentos y nos alegramos mucho por vosotros	1
@Schuma78: ¡¡BRUTAL!! Susanna Griso rebate los argumentos a líder independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 @Albert_Rivera La cara de Susana Griso lo dice todo	1
@Schuma78: ¡¡BRUTAL!! Susanna Griso rebate los argumentos a líder independentista y este la manda a casa a limpiar la ropa y hacer la...	@Schuma78 Buenísimo!!! Están fatal del tarro!!!	1
@PeroniaRepublic: Che, quien es Sietecase? Esto:	@PeroniaRepublic El primer "periodista" en operar por el "suicidio" del fiscal. Y no creo que sea por boludo.	1
@gerardotc: Para quien no sepa francés, le traduzco: "España es un ejemplo para el mundo gracias a que el PP sabe gestionar".	@gerardotc Yo no se francés, pero juraría que no es eso lo que pone.	1
@Syckold: preparando desayuno, creo que no me van a alcanzar los huevos...	@Syckold Si se puede yo te ayudo con gusto estimado	1
@Syckold: preparando desayuno, creo que no me van a alcanzar los huevos...	@Syckold Yo quiero mis huevos que los hagas paya y paca	1
@Syckold: preparando desayuno, creo que no me van a alcanzar los huevos...	@Syckold Aquí tengo unos buenos litros de leche para ti saludos	1
@magalitajes: No es la historia del hombre escondido, es la historia de otra mujer asesinada. BASTA	@magalitajes Que mal elegido ese título por Dios, querer vender una nota en vez que concientizar.	1
@ChileBicampeon: increíble como se descontrola el delincuente Piñera cuando alguien le demuestra que está mintiendo, alz...	@ChileBicampeon @Tirolococomquake Igual no era ilegal lo que hizo, la ley lo permite.	1
@MccArni: Estamos todos de acuerdo en que Soda Stereo es la mejor banda de rock en español de la historia de la humanidad, verdad?	@MccArni Me gustaron todas las repuestas, se nota que hay pasión	1
@MccArni: Estamos todos de acuerdo en que Soda Stereo es la mejor banda de rock en español de la historia de la humanidad, verdad?	@MccArni No, de pop, quizás, pero de rock yo diría Enanitos o Redondos	1
@MccArni Claro que sí, después de los grandiosos HÉROES DEL SILENCIO	@MccArni @toflaqui Uy, lo mío sería: 1. Héroes del Silencio 2. Soda Stereo 3. Café Tacvba 4. Aterciopelados 5. Caifanes	1
@JUSAPOL2018GC: Hoy domingo, propongo un seguimiento MASIVO a la cuenta de @jusapol, que todo el mundo se entere de esta injusticia...	@JUSAPOL2018GC @jusapol totalmente de acuerdo	1
@ErnestoEkaizer: Soy, según este diario, "otro argentino en nuestra vida política". Este es el estilo editorial nacionalista es...	@ErnestoEkaizer @revistamongolia si algo hace falta en este país es españoles como tú, Ernesto. Un abrazo!	1
@viviana_bonilla: Un golpista confeso como Temer escribiendo en diario público ecuatoriano. No solo es vergonzoso para nuestro país sino...	@viviana_bonilla QUÉ VERGÜENZA traer a éste usurpador traidor, bueno entre ratas se huelen los rabos.	1
@MrTristanVidal: Todo el mundo quiere algo	@MrTristanVidal @MundoDTatuajes No es raro!	1
@pccomponentes: SO EAZOI! Comenta con y sigue a PcComponentes y @OMENbyHP y muy importante rellena tus dat...	@pccomponentes @OMENbyHP Lo necesito.	1
@nelsonbocaranda: Rafael Ramirez montó en los pasillos frente al Consejo Seguridad ONU exposición acusando a la oposición de todos los m...	@nelsonbocaranda Este no es más..... porque no ha entrenado lo suficiente.	1
@nelsonbocaranda: Rafael Ramirez montó en los pasillos frente al Consejo Seguridad ONU exposición acusando a la oposición de todos los m...	@nelsonbocaranda Para esto si hay plata! Qué vergüenza!	1
@nelsonbocaranda: Rafael Ramirez montó en los pasillos frente al Consejo Seguridad ONU exposición acusando a la oposición de todos los m...	@nelsonbocaranda Descaro a su máxima potencia la de ese tipejo. No tiene temor a Dios.	1
@carlosblancog: @Diego_Arria ha logrado reunir consensos internacionales fundamentales para la causa democrática venezolana. No se lo pe...	@carlosblancog @barbaradeccs @Diego_Arria @PoleoRafael por ejemplo?	1
@Berlustinho: Dastis tiene las mismas pruebas sobre las injerencias rusas que cuando decía que tenía pruebas sobre la no enseñanza del c...	@Berlustinho La única prueba que tiene el ministro es el gag visionario	1
@IkerCasillas: No me gusta nada verte así! Quiero verte como hasta ahora, como lo que sigues siendo para muchos: una LEYENDA. Orgullosa...	@IkerCasillas Bravo @IkerCasillas Eres uno de los grandes. Un gran gesto.	1
@cristobalsoria: Descubre las 7 diferencias...	@cristobalsoria La diferencia es que el señor de la feria juega con niños a la escopeta falsa y el otro que juega con los niños a los médicos	1
@pccomponentes: SO EAZOI! Comenta con y sigue a PcComponentes y @OMENbyHP y muy importante rellena tus dat...	@pccomponentes @OMENbyHP buena suerte a todos	1
@Soymanorui: Hoy te quiero más que siempre.	@Soymanorui Hoy te quiero más que ayer.	1
@Soymanorui: Hoy te quiero más que siempre.	@Soymanorui Te amo mucho.	1
@Soymanorui: Hoy te quiero más que siempre.	@Soymanorui Hoy te amo más que antes.	1
@Soymanorui: Hoy te quiero más que siempre.	@Soymanorui Cada día te quiero mas que todos los días atrás	1
@Soymanorui: Hoy te quiero más que siempre.	@Soymanorui Hoy te adoro más que nunca.	1
@Soymanorui: Hoy te quiero más que siempre.	@Soymanorui de pies a cabeza	1
@Soymanorui: Hoy te quiero más que siempre.	@Soymanorui Hoy te voy a besar de los pies a la cabezaaa	1

Table A64: Examples of labeled data: Twitter (29)

@ElJuanpaZurita: Bueno pues. Ahora sí. Noches noches noches noches.	@ElJuanpaZurita El que mucho se despidе es porque no se quiere ir	1
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de la s Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky Me alegra mucho. Te lo mereces!	1
@guillier: Prepárense todos, que si la derecha gana será la persecución más brutal, porque el diálogo que han prometido es falso, no van...	@guillier jajajajajaja que penoso, por favor analice bien lo que escribe para no hacer el ridículo.	1
@benshorts: Te sueño cada noche y cada mañana se me rompe el corazón de nuevo.	@benshorts Pense que era el unico a quien le pasaba esto	1
@Societatcc: Vaya con los presos políticos y los fugados, pobrecillos, e staban preparando un golpe de Estado con su ejército y todo.La...	@Societatcc Vaya vaya con la revolución de las sonrisas. El país que pr eparaban no parece el eden de la felicidad....	1
@KRONOS_VI: Yo aprendí a levantarme sin que me dieran la mano.	@KRONOS_VI Excelentes frases	1
@matadoreltiempo: Mercadeando al país	@matadoreltiempo Así o más regalado!!!	1
@thegilbertsosa: Wow los que me amaban hace 10 horas, son los que más están escribiendo cosas feas... pero pues está bien	@thegilbertsosa Supongo que si hubieras dicho algo al respecto de tu desicion esto no hubiera pasado	1
@berrirohemen: Vivimos en un Estado de Derecho y todo el mundo m erece un juicio justo y presunción de inocencia, excepto si a mí me sal e...	@berrirohemen Obviamente tienen derecho a un juicio justo. En un jui cio justo, ellos saldrían condenados y directos...	1
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio Es la estadística de la distribución de la riqueza la que d a vergüenza	1
@luisrodriguezm: El pelotudo de Beto Casella (que tiene como 70 Piru los y se hace el pendejo), ataca a Jorge Lanata y defiende al delin...	@luisrodriguezm Lo dijiste desde el principio, ES UN PELOTUDO!!!!!!	1
@pccomponentes @OMENbyHP xq nunca me toca nada	@pccomponentes @OMENbyHP si no toca uno, toca otro	1
@pccomponentes @OMENbyHP xq nunca me toca nada	@pccomponentes @OMENbyHP Otro mas a ver si este si !!!	1
@vox_es @jusapol Muchas gracias x vuestro apoyo	@vox_es @jusapol Gracias por vuestro apoyo!	1
@vox_es @jusapol Muchas gracias x vuestro apoyo	@vox_es @jusapol Muchas gracias por apoyarnos otro día más	1
@LSisco17: Sólo espero que la RAE acepte "Chavista" como sinónimo d e mamagüevo.	@LSisco17 No vale, la acción y efecto de ser mamagüevo es muy linda para eso.	1
@Yopucheros: Siete años amargando a Carlos Antonio Vélez, ¿cómo no voy a querer a Pekerman!	@Yopucheros Jajajaja. Buenísimo	1
@Bototos: El que dijo que las cosas gratis se valoran menos, no me ha visto en un cóctel.	@Bototos @antobrowne yo si te he visto, que peligro !!	1
@MaihenH: Belkis Solorzano otra víctima del "Socialismo del Siglo XXI" en Venezuela. Perdió su trasplante de riñón por falta de medicame...	@MaihenH Qué tristeza	1
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de la s Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky Felicidades rosa !!! Te lo mereces , tus historias y tu form a de escribir da mucho	1
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de la s Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky Me emociono por ti porque lo tienes bien merecido. Nu nca olvidaré los tiempos que compartimos	1
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de la s Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky @CyM_tw Te lo mereces bravo !!!!	1
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de la s Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky Una gran noticia. Enhorabuena	1
@BrunaHusky: Ahhhhhh!!!! Me acaban de dar el Premio Nacional de la s Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky Muchas felicidades, de una lectora a la que has atrapad o en alguna de tus novelas	1
@auronplay: Por favor que alguien meta en prisión a este señor.	@auronplay Ay que llevarlo al monte del destino y tirarlo	1
@TomiOlava: Yo no quiero ser un opositor irracional pero realmente n o puede un presidente sacar un lateral de esta manera	@TomiOlava Es PRESIDENTE, no es un jugador de fútbol .	1
@TomiOlava: Yo no quiero ser un opositor irracional pero realmente n o puede un presidente sacar un lateral de esta manera	@TomiOlava Por eso es presidente y no futbolista por que sino estaba mos en el orden	1
@Mel_Lezcano: Mis victimas de siempre jajaja @MoliniersReal @MicaVi cicone	@Mel_Lezcano @MoliniersReal @MicaVicicone Que hermosos! Buen tr abajo meli me encanta las foto que les sacas mica esta hermosa	1
@julesgdI: Santo Peje de MacuspanaRefugio de los perdedoresPozo de los rencorososAlborotador de los resentidosVaso de la envidiaTro...	@julesgdI Tu necesitas ir a una misa de sanación a ver si te quitan la e nvidia y la estupidez.	1
@IkerCasillas: No me gusta nada verte así! Quiero verte como hasta ah ora, como lo que sigues siendo para muchos: una LEYENDA. Orgulloso	@IkerCasillas Triste lo de buffon, el sabe que mancho su carrera, que c uando sea viejo lo llamen para recordar	1
@atlanticsurff: Julián Capasso, el productor que insultó a Alfredo Leuco , hoy dijo que sus dichos nada tienen que ver con una ideología...	@atlanticsurff Las redes sociales hacen estragos con las mentiras	1
@lalusiini: cada vez que pienso que soy gilipollas me acuerdo de que h ay gente que tiene como idola a dulceida y se me pasa	@lalusiini No es mi caso pero creo que deberías respetar	1
@miriamnoguerrasM: La normalidad en Catalunya NO es la normalidad del PP	@miriamnoguerrasM Muy a tu pesar, como buena indepe abducida, si nos estamos normalizando.	1
@miriamnoguerrasM: La normalidad en Catalunya NO es la normalidad del PP	@miriamnoguerrasM Claro, la normalidad en una dictadura se llama 15 S .	1
@toppertorpe: La expresidenta del Parlament ha perdido completamen te la cabeza hace tiempo. Se muestra supremacista, racista... Una joya...	@toppertorpe Ha enloquecido. A los españoles no, pero si a la cantida d de extranjeros que tienes en Cataluña..	1
@toppertorpe: La expresidenta del Parlament ha perdido completamen te la cabeza hace tiempo. Se muestra supremacista, racista... Una joya...	@toppertorpe @Jsolamente Usted desconoce lo que es el racismo. Nu estro sistema educativo no sirve de nada si no es efectivo.	1
@El_Intermedio: El videoclip de El Intermedio al ritmo de 'Amigos para siempre' sobre Cataluña y España ...	@El_Intermedio Muy bien, pero estais confundidos, no hay problemas entre españoles y catalanes, me duele que entreis...	1
@AlvaroUnibeVel: Petro creó 50 mil contratos burocráticos en Bogotá, quebró la ciudad y aspira a la Presidencia	@AlvaroUnibeVel Algún día va a proponer y dejar de andar poniendo q uejas trabaje "senador"	1
@DanielRetamozo: Jugó 5 Mundiales. Ganó 19 títulos locales y 3 intern acionales. Es millonario. Es Campeón del Mundo. Tiene 39 años y...	@DanielRetamozo Yo te consuelo!	1
@guruchoirer: ¡Inaudito! Ningún presidente del mundo se queja de lo que se queja Peña Nieto. En todos los países hay redes sociales, soc...	@guruchoirer Pero este es un presidente idiota	1

Table A65: Examples of labeled data: Twitter (30)

@jesusmedinae: A está hora tengo información que se mueven los tentáculos en mi contra, alguien está molestó por que sigo denunciando la...	@jesusmedinae Cuidate mucho hermano, Dios te bendiga... Y continúa tu l	1
@guruchuirer: ¡Inaudito! Ningún presidente del mundo se queja de lo que se queja Peña Nieto. En todos los países hay redes sociales, soc...	@guruchuirer Suenan patético quejándose y queriendo convencer sin razón	1
@guruchuirer: ¡Inaudito! Ningún presidente del mundo se queja de lo que se queja Peña Nieto. En todos los países hay redes sociales, soc...	@guruchuirer @santisolano7 Nadie cree que exista esa actuación, ni por p	1
@2010MisterChip: Después del Suecia 1-0 Italia, esto es lo que ha pasado en las diferentes repescas que se han jugado:1- Honduras 0-0 A...	@2010MisterChip Cierito.	1
@Bollistuff: MIEDO. Te violan y encima se pone en tela de juicio tu compo	@Bollistuff @Arezo No preocuparse, que admita la prueba supone que n	1
@IrantzVarela: Si te estás planteando si las múltiples penetraciones anales	@IrantzVarela Yo verdaderamente creo que ellos piensan que no han hec	1
@BangBangtan_Esp: En unas horas comienza "The late late show con Jame	@BangBangtan_Esp @BTSxAMAs Aún no ha salido y estoy esperando desd	1
@BangBangtan_Esp: En unas horas comienza "The late late show con Jame	@BangBangtan_Esp @BTSxAMAs En ese mismo link puedo ver The ellen sh	1
@BangBangtan_Esp: En unas horas comienza "The late late show con Jame	@BangBangtan_Esp @BTSxAMAs A que hora es?	1
@El Intermedio: "Amigos para siempre", el karaoke de El Intermedio para	@El Intermedio Me encanta la idea ¡Qué pena que se olviden que hay más	1
@xshenaybhi: Hoy estamos, mañana no sabemos.	@xshenaybhi @antonellacald12 Me pasa eso me muerdo que feo me re pe	1
@SilasPicasSocCra: Jajaja cuando estas drogado en Twitter.	@SilasPicasSocCra Que bolido pobrecito...	1
@EmilianoNah: @juanporcino Me quedé encerrado en el baño.	@EmilianoNah @LopezMartinez @juanporcino Con qué le diste a la puert	1
@ElJuanpaZurita: QUE ESTO ES REAL QUE AKGHUIEN LA CONFIRME POR	@ElJuanpaZurita Confirma dónde está el dinero de las donaciones maldita	1
@ElJuanpaZurita: QUE ESTO ES REAL QUE AKGHUIEN LA CONFIRME POR	@ElJuanpaZurita Me muerdo. Ojalá que sea real y certero, por favor, por fav	1
@ElJuanpaZurita: QUE ESTO ES REAL QUE AKGHUIEN LA CONFIRME POR	@ElJuanpaZurita @vodkaustin Es igual de falso que el Tour, a eso le llamo	1
@sugaobseccion: Woo, el aeropuerto estuvo lleno de ARMYs para recibirl	@sugaobseccion Que pena que no tenemos la misma suerte de verlos de c	1
@Davanana: VAYANSE A LA VERGA	@Davanana Pues yo creo que no es la frase, es más la forma de decirlo en	1
@Davanana: VAYANSE A LA VERGA	@Davanana En México también andamos bien progres, usaste a la perfecci	1
@Davanana: VAYANSE A LA VERGA	@Davanana Jajajaja. Ese "vayanse a la verga" te quedó muy mex.	1
@LaGrondonaTwitt: Que alguien le preste una moto a Banega, por el amor	@LaGrondonaTwitt Ya sabemos lo que es capaz de hacer con una moto	1
@Apapell: Repugna la ambigüedad de Coláu tratando de ganarse ahora a l	@Apapell @jmdelaloma La verdad, no sé dónde ven ambigüedad. Más clar	1
@Apapell: Repugna la ambigüedad de Coláu tratando de ganarse ahora a l	@Apapell No es ambigüedad,es tener la cara mas dura que el cemento	1
@Apapell: Repugna la ambigüedad de Coláu tratando de ganarse ahora a l	@Apapell Pues coincido Antonio esa es la palabra.Esa mezcla explosiva de	1
@Cosas_Felices: ¡¡¡Comienza un SuperSorteo!!! Con motivo del estreno	@Cosas_Felices @eccediciones @WarnerBrosSpain Todo un sueño tenerlos	1
@Cosas_Felices: ¡¡¡Comienza un SuperSorteo!!! Con motivo del estreno	@Cosas_Felices @eccediciones @WarnerBrosSpain Seria algun increíble , si	1
@ierrejon: Sólo faltaba Aznar. Cada día más pruebas de que la cúpula del	@ierrejon Y no pasa nada. Este país está anestesiado. No lo entiendo.	1
@AlbertoRavell: Lo que faltaba. Ahora la oposición tiene 2 fracciones parla	@AlbertoRavell Tú bien lo sabías, lo que se venía los que quieren libertad	1
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador.	@elcomerciocom Admirable!	1
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador.	@elcomerciocom @CarlosVerareal El único en esta década que realmente l	1
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador.	@elcomerciocom Esta muy bien el reconocimiento por su trabajo ...	1
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador.	@elcomerciocom Bien merecido. Qué sirva de ejemplo para algunos sinver	1
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador.	@elcomerciocom Un verdadero héroe sin capa , está bien representados a	1
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador.	@elcomerciocom Este caballero debería tener por lo mínimo un salario bás	1
@SDHArgentina: Dale y ayúdanos a seguir difundiendo derechos para hac	@SDHArgentina No. Váyanse a la mierda.	1
@luisvicenseleon: Ligar por un default de Venezuela y pensar que eso sacar	@luisvicenseleon y que los va a sacar según tu ?	1
@luisvicenseleon: Ligar por un default de Venezuela y pensar que eso sacar	@luisvicenseleon Déjese de pendejadas usted	1
@luisvicenseleon: Ligar por un default de Venezuela y pensar que eso sacar	@luisvicenseleon Dices eso con la intención de cambiar la percepción de la	1
@luisvicenseleon: Ligar por un default de Venezuela y pensar que eso sacar	@luisvicenseleon Ok! entonces segun tu mejor es no ligar y que tengan la	1

Table A66: Examples of labeled data: Twitter (31)

@luisvicenteleon: Ligar por un default de Venezuela y pensar que eso sacará al gobierno es una pendejada... al cuadrado.	@luisvicenteleon Estoy de acuerdo	1
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @AngeldebritoOk @FernandezAnibal Se debería haber quedado adentro también!!	1
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal No agradescas tanto!!	1
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal Fuerza Compañero	1
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @omarlavieri @FernandezAnibal Ya llega la visita.	1
RT @Schuma78: ERC en público: "Esto no va de independencia, va de democracia"ERC en privado: "Por la independencia nos pasamos la democracia..."	@Schuma78 Nada nuevo, nada que no sepamos. Son unos Nacionalistas Totalitarios. Así de simple. Este partido debería estar ilegalizado.	1
RT @Schuma78: ERC en público: "Esto no va de independencia, va de democracia"ERC en privado: "Por la independencia nos pasamos la democracia..."	@Schuma78 Y estos son los que tienen miedo de pucherazo?	1
RT @Schuma78: ERC en público: "Esto no va de independencia, va de democracia"ERC en privado: "Por la independencia nos pasamos la democracia..."	@Schuma78 Esto no va de independencia, va de democracia. Esto no va de democracia, va de independencia.	1
RT @Schuma78: ERC en público: "Esto no va de independencia, va de democracia"ERC en privado: "Por la independencia nos pasamos la democracia..."	@Schuma78 Sistema democrático propio de los nacionalistas en Cataluña.	1
RT @sebasantanders: Si Macri, Kuczynski y el presidente de Honduras fuesen de izquierda, la prensa ya estaría haciendo reportajes en extens...	@sebasantanders También hay prensa de izquierda. El problema es que no los lee nadie.	1
RT @JuanMorenoRe: Hemos recusado al nuevo magistrado designado para juzgar la caja B del Partido Popular. Su carrera está estrechamente lig...	@JuanMorenoRe Es de cajón que se rehusa a este juez.	1
RT @JuanMorenoRe: Hemos recusado al nuevo magistrado designado para juzgar la caja B del Partido Popular. Su carrera está estrechamente lig...	@JuanMorenoRe ¿Dudamos? No tengas duda, está de parte de la banda, seguro.	1
RT @JuanMorenoRe: Hemos recusado al nuevo magistrado designado para juzgar la caja B del Partido Popular. Su carrera está estrechamente lig...	@JuanMorenoRe @agaron Recuerdo a un catedrático que me decía: "dudar es no saber". En este caso, no tengo la mínima.	1
RT @publico.es: La UE lanza un proceso contra Polonia sin precedentes por su ataque a la independencia de los jueces https://t.co/ahOSVcFe1e y...	@publico.es Eeeeeeeeeeeeeee!!! UE!!!!!! Y nosotros!!!!!! Que hay sobre la independencia de los jueces españoles?????	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @InesArrimadas No me creo una mierda que voteis todos indepes!!! seguro que tus padres no lo haran	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @InesArrimadas Vota pero no llores cuando estes en el paro	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @InesArrimadas Ni caso a los que quieren imponerte su parecer.	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @InesArrimadas Creo que te equivocas...las elecciones son al Parlament!! No es un referendum ilegal!!!! Cachis!!!!	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @camelianin yo también gadita e independe	1
@IkerCasillas: No me gusta nada verte así. Quiero verte como hasta ahora, como lo que sigues siendo para muchos: una LEYENDA. Orgullosa...	@IkerCasillas Ánimo... amigo...siempre serás grande	1
RT @JCTapiaLMB: Amenazar a los diputados para que voten a favor de la propuesta del Ejecutivo en las designaciones a la CSJ, violando la in...	@JCTapiaLMB Es que esperamos que pase algo distinto en esa votación? Recuerden que estamos ante el órgano más corrupto, un poquito de especias y queda listo el caldo.	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @InesArrimadas VOTA PERO NO LLORES	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @InesArrimadas Ni caso a los que quieren imponerte su parecer. Su último argumento es: "estás en España y eres español te guste o no" y luego ya les da un cortocircuito y les peta el cerebro. Eres grande, los carnegros somos el futuro de Catalunya!!!!	1
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @camelianin yo también gadita e independe	1
RT @JCTapiaLMB: Amenazar a los diputados para que voten a favor de la propuesta del Ejecutivo en las designaciones a la CSJ, violando la in...	@JCTapiaLMB No los amenaza nadie son del mismo chiquero.	1
RT @Schuma78: ERC en público: "Esto no va de independencia, va de democracia"ERC en privado: "Por la independencia nos pasamos la democracia..."	@Schuma78 A la Fiscalía y a la Junta Central Electoral...	1
@lamKatieAngel: Ya estás participando para GANA E uno de estos 5 IPHONES 8 ??? ¿Qué esperas? Correleeee! http://	@lamKatieAngel Claro que sí, me di cuenta por en video y en cuanto lo vi me puse a hacer todo	1
RT @flor.vigna: Que dicen ???	@flor.vigna Obviamente todo te queda perfecto	1
@PabloLanusse: La justicia no es venganza ni morbo. Uno es inocente hasta condena firme. Pero prisión preventiva es válida para garantizar...	@PabloLanusse Tal cual . De acuerdo . Así debe ser .	1
@AntonioJSMazuec: ¿Cómo habéis empezado la semana familia?	@AntonioJSMazuec La he empezado increíble, gracias a tu me gusta soy la más feliz del mundo ahora mismo. Gracias...	1
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal Me gustaría saber cómo hace para usar esta cuenta ud. Desde donde esta.	1
@_anapastor_: La fiscalia belga respondiendo a todas las preguntas...igual que aquí!!!	@_anapastor_ Esto lo tenemos en cuenta? Es lo mismo con la gente de su país, me gustaría...	1
@ghanito: España no tiene salvación posible	@ghanito @SenoritaPuri A mi sin embargo me encanta nuestro espíritu burlón y nuestra capacidad de reírnos de casi todo.	1
@Camila.Cabello: Te amo.....	@Camila.Cabello Te amo más	1
RT @TNFamosos: Mica Viciconte se animó a un primerísimo primer plano sin maquillaje https://t.co/YJAK2Kpe3c	@TNFamosos Es preciosa y natural se muestra tal y como es por eso me encanta	1
RT @IgorMeltxor: ALERTA: nazis de caza por Barcelona aplicando su 155.#L6ARVCat #RepúblicaCatalana #IndependenciaCatalunya #NoPassaran #se...	@IgorMeltxor ese se ha llevado un 155 en toda la cara me encanta, si no fuera un indepencia no le hubiese pasado nada	1
@LuisFlordo: Estamos en EEUU, mañana nos reuniremos con Presidente del TSJ en el exilio y magistrados legítimos ¡atentos!	@LuisFlordo a mi me encanta los EEUU . lastima q ya no puedo ir. Este gobierno me secuestró o los venezolanos todos estamos secuestrados	1
@BeatrizTalegon: La población Catalana tiene una cultura democrática muy por encima de la media española. Por eso aquí no se entiende	@BeatrizTalegon La línea democrática no se...por la intelectual da penica...Por eso aquí pensamos que sois retrasados.	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Hola buenos días	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buenos días, Cómo estás?	1

Table A67: Examples of labeled data: Twitter (32)

@luisrodriguezrm: El pelotudo de Beto Casella (que tiene como 70 Pirulo s y se hace el pendejo), ataca a Jorge Lanata y defiende al delin...	@luisrodriguezrm Dijiste lo que pienso yo también . Exacto	1
RT @DANIROVIRA: ¡Qué gran noticia! 🙌🏻💖 https://t.co/VjgQDFEGKE	@DANIROVIRA Que bien, ojalá que se consiga 🙌🏻💖	1
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices	@marianorecalde Si yo también lo recuerdo siempre que veo sufrir a una	1
la seguridad de los justos.el sufrimiento de los humildes. #Né...	persona humilde lo recuerdo. a El. a su madre, hermana, esposa e hijos...	1
RT @villalobossebas: te amo 🙌🏻	@villalobossebas Yo tambien te amo	1
RT @marley_ok: Nació Mirko!!! 17:26 horas de acá, 51 cm, 3.220kg.Nació	@marley_ok Bienvenido Mirko! Que tengas una larga vida llena de paz, l	1
con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	uz y amor con tu querido papi. Felicitaciones Marley	1
RT @yuyacst: Te amo.	@yuyacst Te quiero linda pequeña...	1
RT @ergasto: Yo estoy a favor de que de una vez por todas la Revista Ba	@ergasto Yo tambien!!!!	1
rcelona se independice del Estado Argentino.		
RT @guillercastello: La provincia remueve a 6 funcionarios involucrados e	@guillercastello @elisacarro Bien hecho	1
n la megaestafa al IOMA que denunciemos junto a @elisacarrohttp...		
@botedeltitanic1: SI ME AUMENTAN EL ABL NO QUIERO VER MAS EXTR	@botedeltitanic1 @ernestoberni @horaciorlarreta Hay cosas buenas tam	1
ANEROS GRATIS EN HOSPITALES PUBLICOS. @HORACIORLARRETA	bién en las plazas, gimnasia para jubilados	1
@elcitizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La di	@elcitizen Yo pienso igual. Eso es lo han demostrado internacionalmente	1
ctadura de Maduro piensa q lo internacional no pasa mas allá...		
@elcitizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La di	@elcitizen Pues yo también así lo veo	1
ctadura de Maduro piensa q lo internacional no pasa mas allá...		
@elcitizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La di	@elcitizen Triste pero yo tambien pienso lo mismo	1
ctadura de Maduro piensa q lo internacional no pasa mas allá...		
@soyFranciscoALV: Primer publicación aquí	@soyFranciscoALV Muchas felicidades. Mucho éxito, saludos	1
RT @dominguezja: SATURNO es la canción más importante en la radi!!	@dominguezja @pabloalboran Si, felicidades Pablete!!! Te lo mereces ♥	1
Felicidades!! @pabloalboran !! Te queremos!! https://t.co/1RWXUgdi		
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no s	@PirryTv Yo no lo creo estoy seguro.	1
on mas que puro teatro,después se sientan a tomar whiskey y a reir...		
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no s	@PirryTv A veces? Jajajaja. Siempre!	1
on mas que puro teatro,después se sientan a tomar whiskey y a reir...		
Hermanos de Panamá y Honduras. @RusoEl23 lo dijo hace semanas y se	@FOXSportsMX @andremarinpuig @RusoEl23 @UpalabraMX Estoy muy	1
cumplió... ¡Al Mundial y Repechaje! (...)	contento por los hermanos Panameños! Esperemos que hagan un gran	1
	mundial!!!	
@CARLOSLGUERRERO Y que dices de Alemania, ellos siguen vivos y Mej	@CARLOSLGUERRERO Bien lo dijiste mucho antes	1
or que nunca, no es karma simplemente es dejar de hacer lo que se teni		
a que hacer.		
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... a	@AlejandroSanz Siempre uso las mías para tocarte la carita cada vez que	1
cariciar, abrazar o dar palmas 🙌🏻 https://t.co/kPw5QN2N8f	te veo y gracias por devolverme siempre caricia con las tuyas. 🙌🏻 te quie	1
RT @carnecrudaradio: Ya que hemos abierto el melón de la República, ¿	@carnecrudaradio Suerte! solo hay que creer firmemente.	1
para cuándo un referéndum para proclamarla en España? #TerceraRepúb		
lica		
@soyFranciscoALV: Primer publicación aquí	@soyFranciscoALV Te deseo mucha suerte y acompañandote para siemp	1
	re saludos	
@soyFranciscoALV: Primer publicación aquí	@soyFranciscoALV Muchísima suerte para lo que viene	1
@caserocosas: QUE NINGUN ORGANISMO DE DDHH.HAYA PLANTADO	@caserocosas @lauala5 Los de los DDHH son como Joan M Serrat que t	1
OPEDUMIO TOTAL POR LA MUE E DE CINCO ARGENTINOS.ES RARISIMO...	odavía no salió a defender a los terro de Barcelona...	1
@BrunaHusky: Ahhhhhhh!!! Me acaban de dar el Premio Nacional de las	@BrunaHusky Ay, cómo me alegro. ¡Cómo me alegro! Enhorabuena y ab	1
Letras... ¡Estoy tan emocionada! Gracias gracias gracias.	razo enorme.	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no s	@PirryTv De acuerdo, no lo dudes. Como crees entonces que todos gana	1
on mas que puro teatro,después se sientan a tomar whiskey y a reir...	n? Cada uno tiene su temporada y así se cuadran!	1
RT @jpedrerol: No teniu vergonya (no tenéis vergüenza). https://t.co/HVC	@jpedrerol @kittlab No tenéis vergüenza.Y así nos podemos tirar toda la	1
30fK9xr	vida... ¿Es lo que quieres? Porque yo no.	1
@ehbildu: Tenemos que construir un muro democrático entre todos los	@ehbildu @ArnaldoOtegi Ya hay muros así para demócratas como vosot	1
demócratas @ArnaldoOtegi	ros.	1
@Pontifex_es: Cristo ha vencido la muerte: Él es nuestra resurrección y n	@Pontifex_es Que así sea. Amén	1
uestra vida. Sean testigos de este mensaje de esperanza.		
RT @1Eme_: Sitios donde se reconoce la República Independiente de Cat	@1Eme_ Si ya esperais en el primer día que todos vayan a "Voy a recon	1
aluña#DuiARV #IndependenciaCatalunya https://t.co/P7iaZND5LA	ocer la independencia de cataluña" así como así	1
RT @AnonymousVene10: Que memoria tan corta tenemos los Venezolan	@AnonymousVene10 @untal_ro No es así. No manipules!!	1
os cada quien siguió con su vida y olvidó a los q ya no tienen la bendici		
ón...		
@soyalannavarro: Algunos de los mejores momentos de la vida son los q	@soyalannavarro Así es es mejor guardarlo a uno mismo	1
ue no puedes contarle a nadie.		
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no s	@PirryTv Es correcto, llevan siglos haciendo lo mismo. Frente Nacional	1
on mas que puro teatro,después se sientan a tomar whiskey y a reir...		
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de esc	@fdbedout jajaja... En serio?! La verdad es que nadie entiende como en	1
ritorio, Colombia y Perú en el min88 tenían que jugar a clasificar a...	3 minutos no arriesgaron la ida al mundial para clasificar a Chile	1
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he vis	@FonsiLoaiza @CatalunyaRadio Gracias por decir la verdad!! 🙌🏻	1
to aquí es de los que llevan banderas de España. Hoy asaltan @Cata...		
@MonederoJC: ¿Cómo es posible que sin presentar ni una sola prueba y	@MonederoJC Algo de verdad habrá o mucha. Al tiempo, es sólo cuesti	1
que el Tribunal de Cuentas diga que la financiación de Podemos está...	ón de tiempo.	1
@oriolguellipui: Pues he estado estudiando el caso un poco y creo que	@oriolguellipui @guillemborrell Suena muy bien como explicación de u	1
sé lo que pasó ayer con la... Lo explico aquí	na posible estrategia	1
@charlieredz: Será porque @VictorAndresGB es íntimo del Presidente de	@charlieredz @VictorAndresGB Vaya vaya interesante, buscaré informaci	1
LAN? Tanto que se ausenta del Congreso para ir al matrimonio del h...	ón: la opinión pública merece saber la verdad y saber quién es quién.	1
RT @Schuma78: ERC en público: "Esto no va de independencia, va de de	@Schuma78 Jajajajaja que raro que digan esto los independentistas ...	1
mocracia"ERC en privado: "Por la independencia nos pasamos la demor		
as...		
RT @jmdelalamo: Hay dos tipos de países: los que no reconocen la inde	@jmdelalamo Cuanta Ridiculez!! Váyanse a vivir a una Isla Desabitada! Y	1
pendencia de Cataluña y los que no saben lo que es Cataluña.	sean Independientes y Felices!! Si viven en es son Españoles Y Punto!	1
RT @GramaticaReal: Atención: la palabra "solo" ya no lleva tilde, nunca.	@GramaticaReal @dennischwartz No estoy de acuerdo con esa regla	1

Table A68: Examples of labeled data: Twitter (33)

RT @marley_ok: Nació Mirko!!! 17:26 horas de acá, 51 cm, 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok Felicidades te mereces lo mejor del mundo	1
@auronplay: Por fin me han puesto los 280 caracteres soy muy feliz en este momento quiero mandar un saludo a mi abuela a ver si un día y...	@auronplay Felicidades por tu cumpleaños y por los 280 caracteres.	1
RT @Blue_Leader_: #DíARVRECONOCIMIENTO INTERNACIONAL CATALUÑA https://t.co/WYQ851rk5W	@Blue_Leader_ Jaja nadie lo va reconocer!	1
RT @luciano_eltirri: Sin palabras!! @cuervotinelli https://t.co/CIN2DmWZMO	@luciano_eltirri @cuervotinelli Que emocionante!!!!	1
@lamKatieAngel: Ya estás participando para GANA E uno de estos 5 IPHONES 8 ??? ¿Qué esperas? Correleeee! http...	@lamKatieAngel si!!! ya estoy participando ojalá ganarme uno eres la mejor te quiero	1
@YouNowLatino: - ¡Mañana a las 22:30h! Tenemos una cita con ellos @rftuertes puch y @BeatrizGH_17 directo en @YouNowLatino http...	@YouNowLatino @rftuertespuch @BeatrizGH_17 Que guay. Mañana nos vemos. Que ganas.	1
@YouNowLatino: - ¡Mañana a las 22:30h! Tenemos una cita con ellos @rftuertes puch y @BeatrizGH_17 directo en @YouNowLatino http...	@YouNowLatino @BeatrizGH_17 @rftuertespuch Que buena noticia	1
@NunkMasKK Hace años vi esos documentales sobre las cárceles de Japón, esta mos años luz de allí...	@NunkMasKK Si, por supuesto. Aca los hacen las vagos de lo que son.	1
@nathcampost: Buenos días! Y feliz día a todas las Bautisters	@nathcampost Feliz día a ti también	1
@itsjoelpimentel: Buenos días	@itsjoelpimentel Buen día cómo estás?	1
RT @marley_ok: Nació Mirko!!! 17:26 horas de acá, 51 cm, 3,220kg.Nació con 37 semanas, necesita un poco aún de cuidados, pero muy felices!...	@marley_ok que alegría por favor!! Felicitaciones les deseo lo mejor	1
@Soymarioruiz: Hoy te quiero más que siempre.	@Soymarioruiz Hoy te amo más que ayer.	1
@CristianSamXXX: Alguno quiere despertarla? Viene con premio	@CristianSamXXX yo quiero y también el premio	1
@JavierBeret: Todo lo que me está pasando me está preparando para lo que quier o ser.	@JavierBeret Ojalá sea así	1
@Mishifucker: Vosotros también le sonreis a los perros por la calle. no?	@Mishifucker @EvaDriobello No es una buena idea. No lo interpretan como un a sonrisa, sino como un gesto hostil	1
@juanbedoya73: Mientras tenga vida y redes, me encargaré de recordarles, que en este país el gobierno pagó a militares por asesinar civil...	@juanbedoya73 @DeCurreaLugo Buena idea para un país sin memoria !!!	1
@MarioBautistaM8: Tu sonrisa te dará un semblante positivo que hará que la gente se sienta mejor a tu alrededor.	@MarioBautistaM8 Tu sonrisa me motiva a hacer varias cosas	1
@Juandiedios: P. Descansa preciosa	@Juandiedios P igualmente amor	1
@luivcenteleon: Lugar por un default de Venezuela y pensar que eso sacará al gobierno es una pendejada...al cuadrado.	@luivcenteleon Venezuela no cae en default porque se ligue, se aupe, se quiere a, igual que ligar a un no ganador.	1
@anaibernal Es evidente lo que el abogado, la "manada" y el juez están pensando: Si es p... se lo merece, vamos a...	@anaibernal Que tristeza y que rabia. Que se puede hacer?	1
@anaibernal Que tristeza y que rabia. Que se puede hacer?	@anaibernal Dime que esto no es verdad, por favor?	1
@unostuitsahi: Esto es chistoso y preocupante a la vez	@unostuitsahi @DanielSamperO Cada día nos sorprenden mas , nos reimos inca nsablemente pero tambien nos preocupan.	1
@bryanmouque: Todos queremos libertad.	@bryanmouque Lastimablemente no todos la tenemos.	1
@Pontifex: es: Cristo ha vencido la muerte. Él es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex es Jesucristo es el camino , la verdad y la vida. Amén	1
@geografosubjeti: En la España democrática no ha parado de haber presos políticos.	@geografosubjeti Pero no es lo mismo, hay que tener en cuenta el hecho diferencial, ser catalán, esto los hace inmunes impunes	1
@cesanoncamilo: Yo estudié gratis, (con beca bicentenario + beca complementaria) y no sabes el compromiso, que sentí estudiando, y ahora...	@cesanoncamilo @sebastianpinera No sabe nada de nada. Solo repite como lor o los guiones que le prepara Goebbels.	1
@gabrielrufian: Nuestra lucha este 21D es por la restitución de las instituciones, gobierno y Presidente legítimo de Catalunya, Carlos P...	@gabrielrufian Claro, claro amigo... pero tranquilízese.	1
@ValuZenere: Es que no puedo parar de ver Stranger Things. Estamos todos igual al no?	@ValuZenere Si Vale te entiendo , esa serie esta increíble!	1
@ziberial: Casi me echan del bar de la risa	@ziberial Muy bueno. Felicitaciones al que lo hizo	1
@carlosmatateal: Lamentable fallecimiento para todos los venezolanos. Freddy Galavis se nos fué esta madrugada. Abrazo y buen viaje quer...	@carlosmatateal Lo siento mucho. Mis mas condolencias para las familias lo recuerdo de la novela Juan la virgen.	1
@XavierGodas: "A los presuntos violadores de Sanfermines, la acusación les pide 22 años. A mi hijo, por una pelea de bar, le piden 50 añ...	@XavierGodas @luis_llach No hay justicia ni nadie que entienda todo esto. Como madre y abuela, lo siento mucho. Ánimos.	1
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes y @elchapuzas Bases sorteo en https...	@pccomponentes @elchapuzas Mucha suerte a todo el mundo y a ver si toca a tope	1
@clarincom: Chubut impulsan un proyecto para que la Policía pueda demorar a encapuchados en protestas.	@clarincom Me parece perfecto	1
@thegilbertrosa: Wow los que me amaban hace 10 horas, son los que más están escribiendo cosas feas ... pero pues está bien	@thegilbertrosa Está muy mal lo que hiciste.	1
RT @sebasantanders: Si Macri, Kuczynski y el presidente de Honduras fuesen de izquierda, la prensa ya estaría haciendo reportajes en extens...	@sebasantanders Estoy de acuerdo contigo, si hasta Maduro a salido de la agenda.	1
RT @PirryTv: A veces creo que todas estas traiciones entre politicos no son mas que puro teatro, después se sientan a tomar whiskey y a reír...	@PirryTv Con estos politicos colombianos, me arrepiento de haber nacido en este país sin oportunidades.	1
RT @AcunaMarcos17: Uno de los días más felices de mi vida! Vamos Argentina	@AcunaMarcos17 Felicitaciones Marcos. Sos el obrero de la selección Argentina de fútbol. Estamos orgullosos de vos.	1
@MarioBautistaM8: Tu sonrisa te dará un semblante positivo que hará que la gente se sienta mejor a tu alrededor.	@MarioBautistaM8 Así es como estar contigo con esta bella sonrisa que solo taverla le robas millones de sonrisas...	1
@ChicasDelCable: Volvemos en Navidad. Pero que quede entre nosotras.	@ChicasDelCable Madre mía de mi vida que no puedo más eh! Me estáis dando unas promos impresionantes! QUE GANAS!!!	1
@primavikings: Hoy más que nunca me repito a mi misma que de todo se aprende, no hay errores en la vida... equivocarnos es vivir... ha...	@primavikings Este mensaje, como cada uno de los tuyos, siempre me levanta el ánimo.	1
@680carlos: Las Universidades huyen del pensamiento crítico. Los alumnos son clientes no estudiantes. Es el negocio de la mediocridad.	@680carlos Totalmente de acuerdo	1
RT @egelt71: El President no acata el cese impuesto por el gobierno de un país extranjero. Justo lo que se esperaba del presidente de mi país.	@egelt71 @Somxserem Y nosotros tampoco.	1
RT @EPN: México no reconocerá la declaración unilateral de independencia de Cataluña. Hacemos votos por una solución política y pacífica.	@EPN Y a nosotros que nos importa si se quieren separar en España. Creo que hay asuntos mas importantes que atender en nuestro país.	1
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo que debe abonarlo para verlo?	@Radio10 Si, claro.	1
@BrunaHusky: Ahhhhh!!!! Me acaban de dar el Premio Nacional de las Letras. ¡Estoy tan emocionada! Gracias gracias gracias	@BrunaHusky No sabe cuánto me alegro. Enhorabuena, querida	1
@kevsho9: Estoy por dar un show en el Luna Park y no lo puedo creer es para un evento privado, por eso no avisé nada, pero es hermoso...	@kevsho9 Ay que genial, me alegro mucho mucho! Felicitaciones por todo	1

Table A69: Examples of labeled data: Twitter (34)



RT @mauriciomacri: No hay discusión. Messi deslumbra siempre. Aplausos para un jugador de otro mundo. 	@mauriciomacri @transportador52 En mi barrio tenemos al E.I. Argentina y urbanizan las 3 villas sin ayuda política A... https://t.co/V0CsCdYc1L	2
RT @soyunapringada_: Maria Rispa haciendo negocio de cobrar a niños thirsty por mencionarles en sus stories creyendo que les hará famosos...	@soyunapringada_ mola, ya cobra más que la mayoría de los que suben zumos detox y demás mierdas.	2
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes + @elchapuzas Bases sorteo en https://t.co/8Tt8eMzskY	@pccomponentes @elchapuzas #AllwaysPcComponentes Un ordenador nuevo siempre se ha tenido	2
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/8Tt8eMzskY	@pccomponentes @elchapuzas #AllwaysPcComponentes Hora de dar el paso a la Next Gen	2
RT @isanseba: Hoy @Albert_Rivera está ganando muchos votos. Me siento más cerca del español harto de pagar impuestos que de los que quieren...	@isanseba @Albert_Rivera ¿Qué tiene que ver el tocino con la velocidad?	2
RT @Lhuchin: Chileno, no te olvides que el CSM que sacó a Bielsa fue Sebastián Piñera.	@Lhuchin El fanatismo político que podre cualquier actividad	2
RT @PabloMM: Es alucinante que Albert Rivera se auto proclame portavoz de la mayoría de los catalanes cuando tiene en Cataluña los mismos A...	@PabloMM Un bluff para quién quiera creerlo.	2
RT @PabloMM: Es alucinante que Albert Rivera se auto proclame portavoz de la mayoría de los catalanes cuando tiene en Cataluña los mismos A...	@PabloMM Están desesperados.	2
RT @Oficialmorris: Un PRI que se multiplicó y tuvo muchos mini PRIs https://t.co/9r8aFbKtbp	@Oficialmorris Un paradigma difícil de superar.	2
RT @antonlosada: Y @marianorajoy hablando como el marido indignado por que su mujer se quiere divorciar, con lo bien que la trata y las cosas...	@antonlosada @AmparoMoLim @marianorajoy Hay divorcio a pesar de AP, partido de Fraga y Mariano que votó en contra... https://t.co/gXWwot5rE	2
RT @xcrstn: si habéis follado con una piba y os correis pero ella no se corre no os la habéis follado... os ha follado ella. Los gillies...	@xcrstn @Navilire comerle el chcho a una chica...	2
RT @rosadiezglez: Extracto del editorial sobre Cataluña del semanario satirico frances "Charlie-Hebdo". https://t.co/aTY8U54VF	@rosadiezglez para tener un like o RT	2
RT @PirryTv: A veces creo que todas estas traiciones entre políticos no son mas que puro teatro, después se sientan a tomar whiskey y a reír...	@PirryTv Un ejemplo clarísimo es clara lopez que de "clara" no tiene mucho	2
RT @PabloMM: Es alucinante que Albert Rivera se auto proclame portavoz de la mayoría de los catalanes cuando tiene en Cataluña los mismos A...	@PabloMM Los ayuntamientos son pantomima de pactos.	2
RT @martindelp: Si eres aficionado a las Chivas y te parece que Lainez es malísimo por ser del América, no te gusta el fútbol, sólo las Chi...	@martindelp realizar un juicio	2
RT @martindelp: Si eres aficionado a las Chivas y te parece que Lainez es malísimo por ser del América, no te gusta el fútbol, sólo las Chi...	@martindelp Sergio Arias fue el mejor portero del mundial 2005 infantil y ni siquiera pudo debutar en primera, son... https://t.co/8Tt8eMzskY	2
RT @martindelp: Si eres aficionado a las Chivas y te parece que Lainez es malísimo por ser del América, no te gusta el fútbol, sólo las Chi...	@martindelp critican a CH14	2
@CARLOSUGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!!... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSUGUERRERO Karma, hermano... eres mexicano	2
@CARLOSUGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!!... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSUGUERRERO su lechita y se me duermen	2
@CARLOSUGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!!... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSUGUERRERO No le hace bien a los estadounidenses comer helados holanda de Chile pork para Chile	2
@CARLOSUGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!!... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSUGUERRERO Los fuertes lloran	2
@CARLOSUGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!!... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSUGUERRERO Se los chupó la bruja	2
@CARLOSUGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!!... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSUGUERRERO El que no se desquita, no es indio.	2
@CARLOSUGUERRERO #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!!... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSUGUERRERO "Por fin habrá quinto (rompo)partido (prexta) antes de que me alburnen!	2
RT @stevenarce: Hola a todos! Abrazo de buenos días ☺ https://t.co/APyyX0seBJ	@stevenarce según los chilenos y argentinos	2
RT @stevenarce: Hola a todos! Abrazo de buenos días ☺ https://t.co/APyyX0seBJ	@stevenarce Ayer se dio todo... Son grandes personas!!!!	2
RT @gabrielrufian: Entiendo que si Rajoy dice que el referéndum se desarticuló el 30/09, el 1/10 se apealó a la gente por puro sadismo.	@gabrielrufian Que cansino...	2
RT @gabrielrufian: Entiendo que si Rajoy dice que el referéndum se desarticuló el 30/09, el 1/10 se apealó a la gente por puro sadismo.	@gabrielrufian Por qué no fuiste a visitar al hospital?	2
RT @Fgarea: Rajoy no ha dicho nada. Al menos nada nuevo: su relato del desafío soberanista, la descalificación del referéndum y defensa de...	@Fgarea Quiero pasar el puente tranquilo.	2
RT @GianThurber: Los ticos son los mejores vecinos que la vida nos pudo dar, celebraron y alentaron a la sele como un panameño más.	@GianThurber Pura vida los solidanos...!!	2
RT @BleuMinette: ¿A ustedes también les pasa que están pensando en alguien cualquiera y DE REPENTE les manda un mensaje? Es como un whatsapp...	@BleuMinette ;500 tuit! https://t.co/v7eLxb7ABlhttps://t.co/QnavXvYtM6	2
RT @Diszain: #MasterChefMx no puedo creer que ganó el equipo rojo 🍷, tendríamos que soportar a doña Julia una semana más. 🙄 https://t.co/UCW6...	@Diszain Faltó esta imagen ¿Quién es? https://t.co/3h5NQ1EwUy	2
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de escritorio, Colombia y Perú en el min88 tenían que jugar a clasificar a...	@fdbedout Por llorar les pasó https://t.co/pprasRhyGk	2
RT @fdbedout: Ahora resulta que con un 3-0 de Brasil y 2 puntos de escritorio, Colombia y Perú en el min88 tenían que jugar a clasificar a...	@fdbedout fue corrupto cuando el chileno le metió el dedo a Cavani	2
RT @EmiRigoni: Muy feliz por la clasificación !! El mejor de todos los tiempos le guste a quien le guste  https://t.co/rMjedUwXZI	@EmiRigoni Messi le pidió una fotografía	2
RT @dlacalle: Mi intervención esta mañana desde Londres en @CNBC desmontando los argumentos independentistas (I) https://t.co/XnhLT338K9	@dlacalle @CNBC Es razonable mejorar situación dentro de la LOFCA	2
RT @Wewerewero: Más que karma es porque ellos se parten el lomo contra Brazil. Italia, Colombia, Argentina, etc. Y Mex... https://t.co/UboAaUmcBz	@Wewerewero Estados Unidos menos del mundo	2
RT @AlbertoRTS1: ULTIMA HORA MUD pedirá a la comunidad internacional más sanciones contra rectoras del CNE y sus familiares por violar de...	@AlbertoRTS1 manda a los borreguitos a votar?	2
RT @Noemi_delaCalle: Que Pablo Iglesias dedique a @Albert_Rivera la primera frase de cuenta de quién está poniendo el sentido común en la...	@Noemi_delaCalle C's es un cáncer para la clase obrera.	2
RT @Noemi_delaCalle: Que Pablo Iglesias dedique a @Albert_Rivera la primera frase de cuenta de quién está poniendo el sentido común en la...	@Noemi_delaCalle @GcMojji @Albert_Rivera Un niño inepito!!!	2

Table A70: Examples of labeled data: Twitter (35)

RT @Noemi_delaCalle: Que Pablo Iglesias dedique a @Albert_Rivera la primera frase da cuenta de quién está poniendo el sentido común en la...	@Noemi_delaCalle @Albert_Rivera Al "coletas" y su banda le quedan dos tel...	2
RT @emilianodiaz34: Orgulloso de ustedes! Son los que dan la cara y suport...	@emilianodiaz34 Vuelves ahora	2
RT @emilianodiaz34: Orgulloso de ustedes! Son los que dan la cara y suport...	@emilianodiaz34 @25_pagantes_ decile a tu viejo y vengán a dirigir Olimpi...	2
RT @emilianodiaz34: Orgulloso de ustedes! Son los que dan la cara y suport...	@emilianodiaz34 la gente no se acuerda que con Ramon le ganamos sólo a...	2
RT @TomasCivetta: Que partido te perdiste Maldonado https://t.co/dRzxx5o...	@TomasCivetta el hermano en el recital	2
RT @TomasCivetta: Que partido te perdiste Maldonado https://t.co/dRzxx5o...	@TomasCivetta Me tienen harta	2
RT @esdonoso_: CTM!!!!lba en una 212 y un abuelito tocaba armónica la zor...	@esdonoso_ Albert Einstein	2
RT @JorgeBustos1: No os fiéis. Rajoy acaba de activar el celeberrimo y nunc...	@JorgeBustos1 A cambio de una reforma constitucional, miedo da!	2
RT @MonsieurSansFoy: El 7-10-34, las tropas de esa II República entraron a...	@MonsieurSansFoy La inteligencia es directamente proporcional... al uso de...	2
RT @MonsieurSansFoy: El 7-10-34, las tropas de esa II República entraron a...	@MonsieurSansFoy 3000 detenidos en un buque-prisión	2
RT @HoracioPagani: Al final a Chile no le sirvió la recurrencia al "escritorio".	@HoracioPagani @HoracioPagani Agradezcan a Messi.	2
RT @TomasCivetta: Que partido te perdiste Maldonado https://t.co/dRzxx5o...	@TomasCivetta Hay que ser pelotudo...	2
RT @JoseLChilavert_: @joaquinfalco La de tu esposa mejor, es mas rica	@JoseLChilavert_ @joaquinfalco grande chilaa idolo	2
RT @JoseLChilavert_: @joaquinfalco La de tu esposa mejor, es mas rica	@JoseLChilavert_ @joaquinfalco no tiene huevos para @ chilal https://t.co/y...	2
RT @JoseLChilavert_: @joaquinfalco La de tu esposa mejor, es mas rica	@JoseLChilavert_ @tincho1387 @joaquinfalco Atendedor profesional de bol...	2
RT @trecet: El OXIMORON PERFECTO: "SOMOS INDEPENDENTISTAS SIN FRONTERAS"	@trecet ahora se van a Murcia que es su otra tierra.	2
RT @fedeposik: Me saqué una foto con mi viejo porque cumple años y salió...	@fedeposik @MaximilianaK tu mamá	2
RT @7LARRA7: Y todo solucionado. https://t.co/Cmp43jFnKG	@7LARRA7 @Frenopatix Te has dejado las Canarias	2
RT @EP_President: 4)Nadie en la Unión Europea va a reconocer esa declarac...	@EP_President @PE_Espana Madre mia como está el patio, he abierto la ne...	2
RT @MariaLujan_Rey: Quiero agradecer a c/u de ls q pensó en los filares d l...	@MariaLujan_Rey Lo que + me indigna.. hubo que ver "asesinar" a 52 inoce...	2
RT @nathcampost: Yo ayer después de ver Coco 🐼💕🧡. https://t.co/25cn...	@nathcampost En qué tienda comercial estuvistes?	2
RT @nathcampost: Yo ayer después de ver Coco 🐼💕🧡. https://t.co/25cn...	@nathcampost No iba a ver video	2
RT @JesusCintora: La crisis catalana tapa las conclusiones de años de investi...	@JesusCintora la Guardia Civil y Policía con sueldos basura.Y millones de fa...	2
RT @JesusCintora: La crisis catalana tapa las conclusiones de años de investi...	@JesusCintora no para todos. he dicho varias veces ¿no han tenido otro mo...	2
RT @abarceloh25: Los senadores del PP jalean a las medidas del 155 mient...	@abarceloh25 No ganamos para tontos	2
RT @abarceloh25: Los senadores del PP jalean a las medidas del 155 mient...	@abarceloh25 El odio les explotará en las narices tarde o temprano!!!!	2
RT @abarceloh25: Los senadores del PP jalean a las medidas del 155 mient...	@abarceloh25 odian catalunya incluso desde cuando no existia el independ...	2
RT @NomasKmail: Echarri no se fue del país.Brancatelli tampoco.Victor Hug...	@NomasKmail En ningún lugar del mundo, van a estar mejor que acá	2
RT @JC_Varela: Felicidades a todos nuestros estudiantes en su día, sigan pre...	@JC_Varela no sea participe de que menos familias compartan momentos ju...	2
RT @JC_Varela: Felicidades a todos nuestros estudiantes en su día, sigan pre...	@JC_Varela Ayude al que esta migracion que no lo dejan pasar a Panama a...	2
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, mu...	@Camila_Cabello @LatinAMAs Nosotros disfrutamos Havana en español	2
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, mu...	@Camila_Cabello @LatinAMAs me enseñás a bailar	2
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, mu...	@Camila_Cabello @LatinAMAs Que ganas de verte ya en directo 🥰 Para cu...	2
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, mu...	@Camila_Cabello @bxiaclavas @LatinAMAs para cuando la colaboración co...	2
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, mu...	@Camila_Cabello @LatinAMAs La presentación fue increíble.	2
RT @Camila_Cabello: @latinamas ayer fue maravilloso estar con ustedes, mu...	@Camila_Cabello @bxiaclavas @LatinAMAs para cuando la foto con fifty ha...	2
RT @davidcc_01: Un abrazo fraternal des de la República Catalana a todos l...	@davidcc_01 Espero que no volváis lloriqueando.	2
RT @Blue_Leader_: #DuARVRRECONOCIMIENTO INTERNACIONAL CATALUÑA https://t.co/WYQB51rkSW	@Blue_Leader_ y El que llora es el pueblo. los que se rien son las élites que...	2
RT @Rajocavi: Vuelvo y digo: Víctor Cantillo es un manual de cómo jugar si...	@Rajocavi @Minutojuniorist Pico el Tony Cross y Cantillo el Sócrates que ju...	2
RT @Rajocavi: Vuelvo y digo: Víctor Cantillo es un manual de cómo jugar si...	@Rajocavi @Minutojuniorist El Sócrates de junior le falta gol pero camina p...	2

Table A71: Examples of labeled data: Twitter (36)

RT @Rajocavi: Vuelvo y digo: Víctor Cantillo es un manual de cómo jugar simple y efectivo. Lo mejor es que lo hace con clase, con categoría...	@Rajocavi es un crack para mí el eje del junior que salda que tiene que pases efectividad casi 100%.	2
RT @planwac: Si se acalla la crítica en una Revolución, se muere la Revolución. El camino es el debate de ideas, la construcción colectiva...	@planwac @floreslegem Los egos nos hace meter la gran pata...lucha, batalla y victoria.	2
RT @planwac: Si se acalla la crítica en una Revolución, se muere la Revolución. El camino es el debate de ideas, la construcción colectiva...	@planwac La oposición desapareció en su errores solo nos afectan protegidos y protectores sordos...Perdimos municipio... https://t.co/3HaF7O8fhH	2
RT @Aracnephobia: 2 semanas ha durado. https://t.co/2ld6nmgQG	@Aracnephobia el video es polemico por? Es un video donde no está atacando a nadie, solo se está quejando de youtube a modo de parodia...	2
RT @Aracnephobia: 2 semanas ha durado. https://t.co/2ld6nmgQG	@Aracnephobia fui a ver el video buscando salseo y resulta que se lo eliminaron por el video defendiendo su hermana y uno de tatuajes...	2
RT @Aracnephobia: 2 semanas ha durado. https://t.co/2ld6nmgQG	@Aracnephobia que pasa con los anuncios o con el tiempo del video? Hay que matarlo por hacer eso o que? Esque joder...	2
RT @pburelli: Los arrestados por ahora son: Nervis Villalobos Cárdenas, Luis Carlos de León, César Rincón Godoy y Rafael Reiter. ¿Los próximos?	@pburelli @Patriotavenex Los diputados del Zulia no están en lista esa sería una sanción magistral.	2
RT @BeatrizCorredor: Hay héroes que se inmolan para salvar a su pueblo y falsos héroes que inmolan a su pueblo para intentar salvarse ellos...	@BeatrizCorredor Aquí se trata de garantizar la impunidad de los criminales que llevan 40 años saqueando Cataluña y... https://t.co/TRiCLJaAR	2
RT @jpedrerol: No tenía vergüenza (no tenéis vergüenza). https://t.co/HvC3OfK9x	@jpedrerol Elecciones el 21 diciembre	2
RT @pardodevera: Hay una manifestación violenta de ultras delante de Catalunya Radio, informan los compañeros/as, algunos encerrados/as dentro...	@pardodevera Cuidado que llevan banderas y dan voces. Apicales el Código Penal peruano anda...	2
RT @pardodevera: Hay una manifestación violenta de ultras delante de Catalunya Radio, informan los compañeros/as, algunos encerrados/as dentro...	@pardodevera Rajoy despertando al fascismo, dándole legitimidad para salir de cacería, lo peor parece que está por... https://t.co/eAnhianR	2
RT @policia: Ayer, hoy y siempre defenderemos la ley, la convivencia y los derechos de todos los españoles. Unidos frente a la adversidad...	@policia (13) Demostrad que tenéis CABEZA y hacéis una HUELGA INDEFINIDA hasta que haya DIALOGO entre las partes... https://t.co/sGrNw3HjXK	2
RT @policia: Ayer, hoy y siempre defenderemos la ley, la convivencia y los derechos de todos los españoles. Unidos frente a la adversidad...	@policia (12) Estoy convencido de que no queréis pasar a la historia como la POLICÍA ANTI-DEMOCRÁTICA Y REPRESIVA... https://t.co/9GkFmuSDS	2
RT @rodrigovega: Que genial que @sebastianpinera vaya a arreglar el Transantiago. Ojalá hubiera sido Presidente antes para tenerlo solucionado...	@rodrigovega @MarioAlexisQ @sebastianpinera sueñen piñeristas...	2
RT @rodrigovega: Que genial que @sebastianpinera vaya a arreglar el Transantiago. Ojalá hubiera sido Presidente antes para tenerlo solucionado...	@rodrigovega @rcorne @sebastianpinera parece que era el portero de la puerta giratoria de la delincuencia	2
RT @rodrigovega: Que genial que @sebastianpinera vaya a arreglar el Transantiago. Ojalá hubiera sido Presidente antes para tenerlo solucionado...	@rodrigovega @sebastianpinera como el puente cau-cau	2
RT @rodrigovega: Que genial que @sebastianpinera vaya a arreglar el Transantiago. Ojalá hubiera sido Presidente antes para tenerlo solucionado...	@rodrigovega @sebastianpinera 🤖 si piñera arreglará el transantiago, van a quedar los buses al revés, o los conduc... https://t.co/vN5GGsRAZ	2
RT @rodrigovega: Que genial que @sebastianpinera vaya a arreglar el Transantiago. Ojalá hubiera sido Presidente antes para tenerlo solucionado...	@rodrigovega @DonAle_CL @sebastianpinera Le cambió los colores 🇨🇱	2
RT @pablocasado: El día de la infamia no quedará impune. Los delincuentes sediciosos acabarán en la cárcel. Cataluña no es independiente y...	@pablocasado, Aplican el 155 de broma. No se atreven	2
RT @pablocasado: El día de la infamia no quedará impune. Los delincuentes sediciosos acabarán en la cárcel. Cataluña no es independiente y...	@pablocasado, Saca tus sucias manos	2
RT @pablocasado: El día de la infamia no quedará impune. Los delincuentes sediciosos acabarán en la cárcel. Cataluña no es independiente y...	@pablocasado, Impune? Gurtel, Castor, Sobres en B, obras, recalificaciónes, Ave, aeropuerto, ... sígo?	2
RT @pablocasado: El día de la infamia no quedará impune. Los delincuentes sediciosos acabarán en la cárcel. Cataluña no es independiente y...	@pablocasado, a cagar extranjero	2
RT @pablocasado: El día de la infamia no quedará impune. Los delincuentes sediciosos acabarán en la cárcel. Cataluña no es independiente y...	@pablocasado, Los independentistas y los peperos bebéis de la misma fuente: cobardía y traición, y en el camino se... https://t.co/Lpuxg73Tx6	2
RT @Santi_ABASCAL: Nos llamen o no nos llamen, el juramento obliga. https://t.co/CWprkYpLO	@Santi_ABASCAL Se tiene que crear una Guardia Nacional para evitar todo esto el gobierno no ha estado a la altura. Somos Españoles no fachas.	2
RT @ilseCeroUno01: Llorando, estoy llorando os lo juro. https://t.co/R36NNWx8J	@ilseCeroUno01 @vicenteprieto El "y?" igual corresponde a que eso lo ha de votar el parlamento europeo, que es el que representa a los ciudadanos, y no su presidente	2
RT @panchocevallosv: Ustedes son valientes y siempre se han entregado a mí máximo para dar alegrías a su hinchada. Confiamos en ustedes! Un f...	@panchocevallosv @CristianBaque3 Quien no espera vencer ya está vencido. "José Joaquín de Olmedo el canto a Junín" vamos ídolo con garra, con fe y fútbol 🇨🇱	2
RT @Ale_Peralta91: Orgulloso de pertenecer a este grupo. @OlimpiaKings semifinalista de la liga sudamericana. https://t.co/jrLHKWwMV	@Ale_Peralta91 @OlimpiaMedia @OlimpiaKings Grande Ale carajo...la rompiste campeón	2
RT @Ale_Peralta91: Orgulloso de pertenecer a este grupo. @OlimpiaKings semifinalista de la liga sudamericana. https://t.co/jrLHKWwMV	@Ale_Peralta91 @OlimpiaKings vos no sos Hijo del Ing. Roberto Peralta ?	2
RT @Ale_Peralta91: Orgulloso de pertenecer a este grupo. @OlimpiaKings semifinalista de la liga sudamericana. https://t.co/jrLHKWwMV	@Ale_Peralta91 @OlimpiaKings Que torneo de mandaste ale, felicitaciones. Vamos olimpia de mi vida.	2
RT @cami_chelmes: Amar a una mina lastimada no es para cualquiera, tenes que saber aguantar y calmar sus miedos, sus inseguridades, si no p...	@cami_chelmes @pichonawesome Pero mujer acá estoy 🤖	2
RT @resistenciaca: Claro que sí!!!! Que oigan a esa gente que no está de acuerdo! 🇨🇱🇨🇱🇨🇱	@resistenciaca Deberíamos empezar con algo un poco más a su altura intelectual.	2
RT @resistenciaca: Claro que sí!!!! Que oigan a esa gente que no está de acuerdo! 🇨🇱🇨🇱🇨🇱	@resistenciaca Muchos están diciendo a las 21:00, mejor todos a una, o se hace a las 21h y a las 22h 🇨🇱🇨🇱🇨🇱	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 un pendejazo	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 El cochino tibetano!!! Me descojono 🇨🇱🇨🇱🇨🇱	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 Ole tus cojones viva el 155 me voy a tomar a tu salud compañero	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 Pedazo retrasado mental. Por eso nos hacemos independientes imbecil, porque escoria como tu no puede... https://t.co/PS0LcO0u5F	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 Ojete, chupame la : https://t.co/H6kzh2PIv8	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 estas hecho un Torrente ...pero de verdad tío.Tienes la gracia de un mono.	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 Quieres ser un cómico...pero te acercas más a retrasado mental Álvaro. Y ya tienes una edad para ser... https://t.co/1GrTOORwJF	2
RT @alvaroorejeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MONGILES !!! VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroorejeda80 Saludos desde la Fiscalía.	2

Table A72: Examples of labeled data: Twitter (37)

RT @alvaroojeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MO NGOLES ¡¡¡ VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroojeda80 El buen meme español. Te recuerdan aún en Gibraltar, vali ente!	2
RT @alvaroojeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MO NGOLES ¡¡¡ VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroojeda80 Y los de TV1 dando imagenes de estas gentuzas. Dimision es en TVE ya mismo.	2
RT @alvaroojeda80: A LA PUTA CARCEL CON TODOS LOS GOLPISTAS MO NGOLES ¡¡¡ VIVA ESPAÑA !!! https://t.co/qU52xndwFJ	@alvaroojeda80 Pobre imbecil	2
RT @ensanro: La Declaración Unilateral Independencia es un error politico que hará retroceder a #Catalunya y fortalecerá a la derecha españ...	@ensanro ¿ quedan derechos por perder ?	2
RT @ensanro: La Declaración Unilateral Independencia es un error politico que hará retroceder a #Catalunya y fortalecerá a la derecha españ...	@ensanro Si la "izquierda española" hubiese pasado cuentas...	2
RT @ensanro: La Declaración Unilateral Independencia es un error politico que hará retroceder a #Catalunya y fortalecerá a la derecha españ...	@ensanro @Caninator Lo de hoy es la demostración de dos gobiernos fac has excusándose con un enemigo externo. Mientr... https://t.co/2HxvAKSA	2
RT @ensanro: La Declaración Unilateral Independencia es un error politico que hará retroceder a #Catalunya y fortalecerá a la derecha españ...	@ensanro @Caninator solo me pregunto.. y encima les entregamos los arg umentos.. por ser más reflexivos y analíticos... https://t.co/OKFvmoGjh	2
RT @Elaguujon_: A ver lo que hace mañana este gilipollas sin DNI. https://t.co/st1NM7nZ6	@Elaguujon_ @isitenos4 Por culpa de tontos de los cojones como este me mo, esta Cataluña como está, esperemos que le... https://t.co/7jRZl3qvCz	2
RT @albertmercade: #Alerta! Manifestació violenta a @CatalunyaRadio ... https://t.co/afsw2Ae7sz	@albertmercade @gerardquintana @CatalunyaRadio Vaya, una pandilla de los del diálogo intentando entrar a repartir t... https://t.co/JHmDcUreH	2
RT @ariasborque: Un policía local retira la bandera de España en Pineda d el Mar (Barcelona). El alcalde es Xavier Amor (PSC) #Independencia...	@ariasborque Que cambie la profesión de policía por la de payaso. De na da.	2
RT @ariasborque: Un policía local retira la bandera de España en Pineda d el Mar (Barcelona). El alcalde es Xavier Amor (PSC) #Independencia...	@ariasborque Xavier pon la bandera de España y expedienta al Policía.	2
RT @ariasborque: Un policía local retira la bandera de España en Pineda d el Mar (Barcelona). El alcalde es Xavier Amor (PSC) #Independencia...	@ariasborque Detención del alcalde y cualquier otro implicado. Disolución de la Corporación	2
RT @ariasborque: Un policía local retira la bandera de España en Pineda d el Mar (Barcelona). El alcalde es Xavier Amor (PSC) #Independencia...	@ariasborque "Ostia puta, el Jordi está llorant"...Hay mucho trabajo que ha cer en la nueva República	2
RT @ariasborque: Un policía local retira la bandera de España en Pineda d el Mar (Barcelona). El alcalde es Xavier Amor (PSC) #Independencia...	@ariasborque Le quedan dos telediarios ...	2
RT @ariasborque: Un policía local retira la bandera de España en Pineda d el Mar (Barcelona). El alcalde es Xavier Amor (PSC) #Independencia...	@ariasborque Espero que este cabron este expulsado del cuerpo por traid or cobarde y basura	2
@tonialba Preneu-vos ja la pildora vermella.	@tonialba 21 de diciembre, habrá elecciones	2
@tonialba Preneu-vos ja la pildora vermella.	@tonialba La resaca va a ser de Puta madre. No os vais a levantar en días. Solo para ir al psiquiatra. Menudos iluminados.	2
RT @JosPastr: Los de Junts pel Sí y la CUP no pueden presentarse a las ele cciones del 21 si quieren ser coherentes. Pero eso es suponer dem...	@JosPastr Puigdemont, Junqueras y Trapero CESADOS,a la puta calle, a tra bajar de una puta vez y ahora que lloren un... https://t.co/H3KR7ssg30	2
RT @CervantesFAQs: Ni una sola crítica de @sanchezcastejon al PP. A ver s i os dais cuenta ya de una vez. #DuiARV	@CervantesFAQs @sanchezcastejon Lo gracioso es que no se queden a vot ar NO. Los únicos que se lo han dicho a los i... https://t.co/iSGr2MXP8G	2
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el gol pe, se aplica el 155 CE y el Estado garantizará elecciones democráti...	@Albert_Rivera No se dan las condiciones para unas elecciones en #Liberta dHay mucha mi... https://t.co/s32KWWjw7C	2
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el gol pe, se aplica el 155 CE y el Estado garantizará elecciones democráti...	@Albert_Rivera Las prisas son malas consejeras...quiero equivocarme.	2
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el gol pe, se aplica el 155 CE y el Estado garantizará elecciones democráti...	@Albert_Rivera Pobrecillo, cuentas para no dormir.	2
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el gol pe, se aplica el 155 CE y el Estado garantizará elecciones democráti...	@Albert_Rivera entendemos como un 'os moleremos a palos', no?	2
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el gol pe, se aplica el 155 CE y el Estado garantizará elecciones democráti...	@Albert_Rivera Rivera se te ve el plumero	2
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el gol pe, se aplica el 155 CE y el Estado garantizará elecciones democráti...	@Albert_Rivera @CiudadanosCs Un presidente de un país no puede convo car elecciones en otro país vecino.	2
RT @Albert_Rivera: Finalmente como proponíamos en Cs. para parar el gol pe, se aplica el 155 CE y el Estado garantizará elecciones democráti...	@Albert_Rivera Ahora a por todas, el partido de la verdadera democracia, @InesArrimadas president, vivo en graná y no quiero pasaport	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero Definitiva y lamentablemente, no conoce la realidad. El resulta do será INCONTROLABLE	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero Ya veremos si habrá garantías!!!!	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero En alguien tiene que estar la inteligencia y ese tio la tiene	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero Cuanto menos dure el 155, mejor verdad? dificultades en pon erlo a la práctica?	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero @emocionycambio hace mucho que dejé de creer en las punt adas sin hilo...	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero Por decir algo... como puede disolver un parlame en el que y a no pinta nada?	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero Ooops yo tengo pedicura y no me va a ir	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero Quien es Rajoy?	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero Ahora ya si eso miramos la gurtel?	2
RT @julia_otero: 21 diciembre, elecciones en #Cataluña. Rajoy ha sorprende do a todos. Esto no figuraba en las previsiones de analista algun...	@julia_otero ¿Dónde está la trampa?	2
RT @CancinoAbog: Cuell jnuna ? Descríe asesino? Delincuente de lesa hum anidad? Usted lo es! https://t.co/vUkxyxKcm	@CancinoAbog @LuciaEbaena Esos apelativos "tan bien ganados"...no se l os quita nadie.	2
RT @edufeioik: El parlamento se limpia un poco https://t.co/vZ2sfD8QBd	@edufeioik Lindos nombres	2
RT @edufeioik: El parlamento se limpia un poco https://t.co/vZ2sfD8QBd	@edufeioik @ingluis451 Por rsto soy capaz el 10/12.de hacer un Ayuno Cel ebratorio!	2
RT @edufeioik: El parlamento se limpia un poco https://t.co/vZ2sfD8QBd	@edufeioik Excelente: en el Congreso debemos tener legisladores que pien sen en el bienestar de la Nación y no solam... https://t.co/NWnCsEP46A	2

Table A73: Examples of labeled data: Twitter (38)

RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio mas tonto que pegarle a un padre	2
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio ¿Los coches de la guardia civil que destruyeron? Deja ya de hacer el ridículo	2
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio España responde al nacionalismo con democracia, elecciones constitucionales	2
RT @TNFamosos: Mica Viciconte se animó a un primerísimo primer plano en maquillaje https://t.co/YAK2Kpe3c	@TNFamosos @MicaViciconte Soy insorpotable. Lo único que sabes es pellear.	2
RT @juancruzgandolf: Yo y mi costumbre de querer a las personas mas de lo que me quieren a mi	@juancruzgandolf Somos dos...	2
RT @AntonioMaestre: Los fascistas hoy saldrán de caza en Cataluña. Que se oculte la buena gente porque serán tiempos complicados.	@AntonioMaestre Eres un verdadero sinvergüenza, una vergüenza como persona	2
RT @luipardo: Feijóo manda un mensaje unionista a Cataluña. Lo que pasó después te sorprenderá... https://t.co/iXdGJ2V8RV	@luipardo @CarodRoviraJLI Prueba su propia medicina	2
RT @Pablo_Iglesias_: La DUI no es legítima, pero la cuestión es: ¿por qué un 48% de catalanes quiere irse de España? La respuesta está en e...	@Pablo_Iglesias_ Tú a tu pedo	2
RT @Pablo_Iglesias_: La DUI no es legítima, pero la cuestión es: ¿por qué un 48% de catalanes quiere irse de España? La respuesta está en e...	@Pablo_Iglesias_ Falso y pedante vete a Venezuela ya	2
RT @Pablo_Iglesias_: La DUI no es legítima, pero la cuestión es: ¿por qué un 48% de catalanes quiere irse de España? La respuesta está en e...	@Pablo_Iglesias_ Sigues dormido.	2
RT @IgnatiusFarray: Las élites de la burguesía catalana ya tienen lo que quieren: su propio Estado para montar su empresa, igual que el PP...	@IgnatiusFarray Para ser elite no tengo un puto duro!	2
RT @IgnatiusFarray: Las élites de la burguesía catalana ya tienen lo que quieren: su propio Estado para montar su empresa, igual que el PP...	@IgnatiusFarray Se te da mejor la commedia, que el analisis politico	2
RT @Cumparini: Me topé a Kast saliendo de Santiago por el finde largo https://t.co/kNL7depNON	@Cumparini @liliana_edgar Te equivocaste eran algun comunista burgués aplicando odio	2
RT @Cumparini: Me topé a Kast saliendo de Santiago por el finde largo https://t.co/kNL7depNON	@Cumparini Por los comentarios, heriste los sentimientos de varios #tchustchis	2
RT @Cumparini: Me topé a Kast saliendo de Santiago por el finde largo https://t.co/kNL7depNON	@Cumparini de verdad que me estoy desayunando contigo	2
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde a la que te pario tuerto hijo de una remil putas.	2
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde HOY SE CUMPLEN 7 AÑOS DE UN CHORRO MENOS PARA EL PAÍS.	2
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde Serán recordados como los ladrones más grandes del país, hay que ser chorro para sacarle el puesto a Menem	2
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde Hay conga en el infierno.	2
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde Pegue el salto Marianito...usted es joven, antes de que se fuera demasiado tarde...	2
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde cuanto emoción!!! El tuerto sinvergüenza.	2
RT @marianorecalde: Te recordamos siempre junto a la risa de los felices, la seguridad de los justos, el sufrimiento de los humildes. #Né...	@marianorecalde Entré para ver los comentarios a favor de los seguidores por el #NestorVive...siguen solo troll... https://t.co/uDWi3Ex42V	2
RT @Elaguijon_: Por ahora han reconocido la independencia de Cataluña Otegi, el PANCETAS, los argentinos de Podemos y un banquero de Andorra...	@Elaguijon_ @JuliaRa51981158 2018 fin del secreto bancario	2
RT @Elaguijon_: Por ahora han reconocido la independencia de Cataluña Otegi, el PANCETAS, los argentinos de Podemos y un banquero de Andorra...	@Elaguijon_ El mundo entero es facha. El siguiente paso es ...	2
RT @mmendizabal1: Votar en secreto no es de valientes	@mmendizabal1 @SSantiagoSegura Mucho mejor salir de la sala para no tener que mojarse?	2
RT @mmendizabal1: Votar en secreto no es de valientes	@mmendizabal1 Que interesada y partidista eres...algo ha quedado en Catalunya la picaresca española.	2
RT @AndresFelipe: ¿O sea que Santrich, culpable de asesinato de mi madre, me puede denunciar por decirle asesino? Como va el país el tribuna...	@AndresFelipe @akenatton3 A esta crápula la deben meter a la cárcel para que pague por sus crímenes y no se sienta igualado #SantrichAsesino	2
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @CatalunyaRadio Sales poco entonces!!!	2
RT @PeriodicoWarca: Berbatov, excompañero de CR7: "Cristiano es brillante, pero Messi es el puto amo, el mejor jugador de la historia" http://	@PeriodicoWarca Vives obsesionado...preocupate por ti macho	2
RT @auronplay: Fue irse Neymar del Barça y todo mal, huracanes por todas partes, terremotos, España rompiéndose, incendios... puto Bartomeu...	@auronplay pienso que te has mojado poquito.	2
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @revueltageneral @CatalunyaRadio Dame veneno que quiero morir dame veneno	2
RT @FonsiLoaiza: Llevo 4 años en Catalunya, la única violencia que he visto aquí es de los que llevan banderas de España. Hoy asaltan @Cata...	@FonsiLoaiza @MariaPastelles @CatalunyaRadio Cambia las gafas porque tienes un problema ayer se vio los estudiantes	2
RT @luiz_sisepuede: Enhorabuena ya no tendréis que elegir entre la papeleta azul, la naranja y la roja para votar a la derecha #L6ARVCat https://t.co/...	@luiz_sisepuede como vosotros con JxSí, la derecha rancia y corrupta anteriormente conocida como Pujol S.L.	2
RT @luiz_sisepuede: Enhorabuena ya no tendréis que elegir entre la papeleta azul, la naranja y la roja para votar a la derecha #L6ARVCat https://t.co/...	@luiz_sisepuede vayan unidos y cantaremos pscoe y pp la misma mierda esta	2
RT @Diego_Arria: A MUD(lodo)se le aplica Génesis 3:19 "Polvo eres, y en polvo te convertirás"	@Diego_Arria profeta del desastre	2
RT @Eanisteguieta: Quienes han venido jugando a mantener a Maduro hasta el 2019, mientras aparentemente serían oposición, son mas culpables que...	@Eanisteguieta @wizardsf Ramos Allup, Timoteo, Falcon, Rosales, Borges y otros detrás de bastidores.	2
RT @callmeflors: dudo que haya muestra de amor mas grande que acceder a dar con alguien sabiendo que al otro día hay que levantarse re...	@callmeflors después de tu cumpleaños cuando me acompañaste a trabajar @YaninaSantiago1	2
RT @MariaFdaCabal: Ahora los mamertos dirán que la culpa es del gobierno, que el País estaba "deprimido". Que necesitaba otro viaje a la H...	@MariaFdaCabal Estudié mejor, solo la amestrazaron a decir Mamerto en esa escuela de mascotas del CD. Sepa argumentar	2
RT @MariaFdaCabal: Ahora los mamertos dirán que la culpa es del gobierno, que el País estaba "deprimido". Que necesitaba otro viaje a la H...	@MariaFdaCabal @CeDemocratico Cizañeros son responsables que la paz no avance. El pueblo les va a cobrar electoralmente votando por candidato que fortalezca la Paz.	2
RT @NataliaPastor: El arquitecto Oscar Tusquets, ayer en Madrid: "No nos abandonéis". Se te pone la piel de gallina. https://t.co/j0kinMk4	@NataliaPastor @alfonso_ussia Todos a la manifestación en Barcelona este domingo en contra del separatismo!!!	2

Table A74: Examples of labeled data: Twitter (39)

RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo no vayan a votar por el PRIAN eso sería autojutiarnos, voten po... https://t.co/FVZMeGk5/n	2
RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @jrisko En unos días se llegó al precio... Lástima México	2
RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo De qué manera lo amenazarían (personal/familiar) para que declinara ante semejante injusticia y corrupción. Pero no es Venezuela	2
RT @esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoción. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @jrisko TANTA CORRUPCIÓN en el MALEADO SISTEMA. #MiVotoParaAMLO y con #MORENA	2
RT @MatiRomeRP: Este si que era un juego la puta madreeeee https://t.co/Uvhr5l81tR	@MatiRomeRP @Arielg9416 Mas trauma me acuerdo!	2
RT @kevsho9: VIDEO NUEVO! Terminamos el mes de Halloween en el canal con un sketch y estoy acompañado https://t.co/iRLPIMtDdf #Halo	@kevsho9 YA TENGO TODAS LAS LETRAS PERO NO ME RECONOCE EL LINK AYUDA.QUE HAGO @kevsho9	2
RT @kevsho9: VIDEO NUEVO! Terminamos el mes de Halloween en el canal con un sketch y estoy acompañado https://t.co/iRLPIMtDdf #Halo	@kevsho9 Llegué tarde pero quería decirles QUE AMO CON MI ALMA LOS ESPECIALES QUE HACEN ENTRE TODOS HAGAN MAS AHRE	2
RT @rubencaravaca: ¿Nadie dice nada de la aprobación del CETA en el Senado? ¿un golpe de estado real contra nuestra soberanía? ¿Qué nos está...	@rubencaravaca No ven con la bandera!	2
RT @IgorMeltxor: ALERTA: nazis de caza por Barcelona aplicando su 155. #6ARVCat #RepúblicaCatalana #IndependenciaCatalunya #NoPassaran #se...	@IgorMeltxor Los extremos se atraen, de vuestros polvos estas ...	2
RT @IgorMeltxor: ALERTA: nazis de caza por Barcelona aplicando su 155. #6ARVCat #RepúblicaCatalana #IndependenciaCatalunya #NoPassaran #se...	@IgorMeltxor Que esperais ahora? Afrontarlo, construir...	2
RT @ladivaza: rt o tendrás 1000 años de mala suerte https://t.co/57jxyOvLQX	@ladivaza Mi teléfono esta peor	2
RT @Angelalerena: Justicia viene perdiendo por goleada, pero de vez en cuando levanta la cabeza, sale jugando desde el fondo y nos ilusiona...	@Angelalerena Lastima que sólo creas que esto es sólo lo bueno de la justicia...	2
RT @BibiMonteroDR: Cómo es posible que en el extranjero si hablen de la corrupción en México y aquí las malditas televisoras callan! #YaBas...	@BibiMonteroDR @GarcaMario La gente se la pasa viendo telenovelas y le embrutece el cerebro.	2
RT @renfrancis: Ya casi es Noviembre y yo sigo sin ti.	@renfrancis Porque vas a estar mas feliz	2
RT @elephantiny: Si no gritaste cuando apareció Micaela con este peinado NO SE QUÉ HACES EN ESTE FANDOM #HalloweenYoutuber https://t.co/mX6...	@elephantiny @MicaSuarez12 avos te aparecion bien el URL o te decia que estaba mal?	2
RT @agpuayas: Nuestro Presidente Vitalicio del Movimiento @35PAIS, @MashiRafael, envia un mensaje a toda la militancia del #Guayas. @marce...	@agpuayas @35PAIS @MashiRafael @marcelaguinaga Es muy fácil, tome un vuelo derecho a Ecuador	2
RT @CUBONoticiasMX: ¿Mira @alfredodelmazo así "tratan" a los reos del Penal Neza Bordo. ¿Algo qué decir? https://t.co/5MjyNwYKvF	@CUBONoticiasMX @alfredodelmazo @brozoxmiswebs ausencia de autoridad a todos los niveles...	2
RT @lopezobrador_: Claudio X. González, quien destapó a Meade y siempre me ataca, es un salinista megacorrupto que participó en el fraude e...	@lopezobrador_ Es tiempo de recuperar la República.	2
RT @lopezobrador_: Claudio X. González, quien destapó a Meade y siempre me ataca, es un salinista megacorrupto que participó en el fraude e...	@lopezobrador_ Manolito no es corrupto? Nunca lo ha sido?	2
RT @jorgeramosnews: O sea, en México no pasa nada. Hay una clase política casi intocable, donde se cuidan unos a otros, y urge cambio tipo...	@jorgeramosnews Corruptos Jorge Ramos y Univision	2
RT @jorgeramosnews: O sea, en México no pasa nada. Hay una clase política casi intocable, donde se cuidan unos a otros, y urge cambio tipo...	@jorgeramosnews @GaelGarciaB A quien quieren, tanto Frederic que Jorge no quieren Obrador.	2
RT @ElComentarioEs: HISTORIA DE ESPAÑA https://t.co/5MjyNwYKvF	@ElComentarioEs QUE RISA!!   	2
RT @Rubius_: Youtube va un poco lento hoy  No se procesa bien el video	@Rubius_ calla estafador estafas a tu gente con tu simple personaje te portaste mal en peru	2
RT @carlosmatareal: Lamentable fallecimiento para todos los venezolanos. Freddy Galavis se nos fué esta madrugada. Abrazo y buen viaje quer...	@carlosmatareal Desde el cielo, gloria eterna para este gran señor.	2
RT @GuillermoBotero: @BancoRepublica reduce tasa de interés en 25 pb a 5%. Están muertos del susto con desempeño económico. Ojalá bancos l...	@GuillermoBotero @JuanLozano_R @BancoRepublica Muertos estamos los colombianos con el ministro cardenas	2
RT @GuillermoBotero: @BancoRepublica reduce tasa de interés en 25 pb a 5%. Están muertos del susto con desempeño económico. Ojalá bancos l...	@GuillermoBotero @amparomeza @BancoRepublica El traslado que haría l a diferencia sería de recursos de las arcas del gobierno a los bolsillos de los consumidores y emprendedores.	2
RT @AlvaroUnibeVel: Retén terrorista, Farc en el Cauca, guardaron el uniforme y sacaron las armas que se supone habían entregado https://t...	@AlvaroUnibeVel la falsa PAZ de ese presidente Santos	2
RT @GuillermoBotero: Primero suspendieron fumigación aérea y ahora quieren despenalizar siembra de 3.8 hectareas de cultivos ilícitos. Dond...	@GuillermoBotero Todavía habrá colombianos que crean que la tal paz de Santos era para traernos progreso y no para inundar al país de coca?	2
RT @GuillermoBotero: Primero suspendieron fumigación aérea y ahora quieren despenalizar siembra de 3.8 hectareas de cultivos ilícitos. Dond...	@GuillermoBotero El día que elegimos a Santos.	2
RT @SF9official: ¿Esto es todo? https://t.co/MFZHWjvQxZ	@SF9official Esta en español de verdad????????	2
RT @SF9official: ¿Esto es todo? https://t.co/MFZHWjvQxZ	@SF9official Me estoy enamorando	2
RT @SF9official: ¿Esto es todo? https://t.co/MFZHWjvQxZ	@SF9official ¿Como podes ser tan lindo?	2
RT @LoyalToPimentel: Imaginense ser la novia de joel y boludear, que te maquille, planche el pelo, etc aaa yo quiero https://t.co/Kx7k736lKb	@LoyalToPimentel De donde es eso?	2
RT @gerardot: Esta noche en CAT hay nazis atacando, pero llamarlos nazis sería un lío pq ya habíamos quedado en que los nazis eran los de...	@gerardot: Tendrá que empezar a tomar el control la policía nacional.	2
RT @Pablo_Iglesias_: 3. Seguiremos defendiendo el diálogo y un referéndum legal y pactado como mejor solución a la crisis en Catalunya	@Pablo_Iglesias_ Hasta el monopoly te viene grande.	2
RT @Pablo_Iglesias_: 3. Seguiremos defendiendo el diálogo y un referéndum legal y pactado como mejor solución a la crisis en Catalunya	@Pablo_Iglesias_ Consideracion en tres palabras a las 3 tuits	2
RT @VillegasPoljak: Gaceta OficialRamón Lobo Pte BCVWilmar Castro VP(E) EconomíaSimón Zerpa Min(E) FinanzasLuis Soteldo Pte INT#Vene...	@VillegasPoljak Zapatero a su zapato.	2
RT @VillegasPoljak: Gaceta OficialRamón Lobo Pte BCVWilmar Castro VP(E) EconomíaSimón Zerpa Min(E) FinanzasLuis Soteldo Pte INT#Vene...	@VillegasPoljak La situación económica no se soporta.	2
RT @padrezurdo: @laidygoomez Estas buena y es lo unico q te salva. Debes vender tu cuerpo y no tu alma. Arrastrada! Bonita por fuera mas...	@padrezurdo @laidygoomez Chavista resentido ? O falta de cariño ?	2
RT @jtgazmuri: ¿ O los ASESORES que ayudaron a la Josefa a desfondar el Municipio de Providencia?	@jtgazmuri Y pensiones fabulosas, pero solo para los compañeros.	2

Table A75: Examples of labeled data: Twitter (40)

@jtgazmuri : O los ASESORES que ayudaron a la Josefa a desfondar el M unicipio de Providencia?	@jtgazmuri Cálculos realizados por Osvaldo Andrade y su señora	2
RT @PiensaPrensa: [LO ULTIMO] PIÑERA: "Si la Presidenta cierra Punta Pe uco, voy a volver a abrirla". https://t.co/e7lau5Hbdn	@PiensaPrensa Por qué se les debe un trato especial?	2
RT @ziberial: Casi me echan del bar de la risa https://t.co/R7t39uKN8K	@ziberial Premio Ganador indiscutible!!	2
RT @ziberial: Casi me echan del bar de la risa https://t.co/R7t39uKN8K	@ziberial @ElCipayo Un poco más de respeto con Richard Wagner y sus walkirias.	2
RT @ziberial: Casi me echan del bar de la risa https://t.co/R7t39uKN8K	@ziberial Es tremendo, hizo el mejor spot de campaña de su vida.	2
RT @rafaellacava10: Carro d Dracula en los alrededores d la Chet y conse guimos a unos delincuentes q tenían una casa para guardar lo q se f...	@rafaellacava10 @RICARDOPLAZA69 pena de muerte es que se merecen esos bichos.	2
RT @dvlzura: RT si leiste "una nota gratis" https://t.co/dEFc5sabAd	@dvlzura tal vez tenga mala la vista	2
RT @ElPadrastroXXX: --Me gusta ir a visitar a mi hermana y encontrarme a mi Cuñado solo... https://t.co/FKnS59ZlGj	@ElPadrastroXXX Jee se quita el pobre chavito pasivo delgadito	2
RT @conderigoli: De pronto toda la oposición Ni MU de Maldonado... ni p ara disimular el asqueroso uso político q hicieron CADA DÍA A CADA IN...	@conderigoli @argprik No dejes de ver esta entrevista brillante https://t co/Auz9x81V5F	2
@dosogas actuo mejor yo q ustedes ahre	@dosogas la cara de el ketchup	2
@dosogas actuo mejor yo q ustedes ahre	@dosogas Tipico de YouTube xD https://t.co/Hbxpr615c	2
RT @eltanomdp: No olvidar a estos mercenarios oficialistas son cómplices de una desaparición forzada seguida de asesinato #LoAsesinaron htt...	@eltanomdp @mononaban Por bananas los monos bailan	2
RT @eltanomdp: No olvidar a estos mercenarios oficialistas son cómplices de una desaparición forzada seguida de asesinato #LoAsesinaron htt...	@eltanomdp Más rompe pelota están con esto..	2
@JavierMMoure @jinsua49 Con esa vocesita de nenita, hacé shhhhh y no jodas más. https://t.co/pRoROJKlK3	@JavierMMoure Y se quedan cortos	2
@JavierMMoure @jinsua49 Con esa vocesita de nenita, hacé shhhhh y no jodas más. https://t.co/pRoROJKlK3	@JavierMMoure @Gorniz70 Si no te gusta arma un partido y gana las elec ciones !!	2
RT @GonzaGravano: Y YouTube es cuestión de tiempo, entiéndanme que quiero hacer algo distinto y no el contenido que hacen todos. Paciencia...	@GonzaGravano Yo te estoy dando más visitas porque con los que tenes los repito mil veces	2
RT @ladivaza: rt o tendrás 1000 años de mala suerte https://t.co/S7jxyOVL OX	@ladivaza El spam ahí rico	2
RT @ladivaza: rt o tendrás 1000 años de mala suerte https://t.co/S7jxyOVL OX	@ladivaza @AriadnaGold como siento que esta mi celular en este momen to	2
RT @hernanpablo: Pasamos de un gobierno que se peleaba con Paul Sing er en tribunales de NY, a uno que se pelea con la familia de un desapac...	@hernanpablo ¿Cuándo fue que se jodió la Argentina?	2
RT @Camus_Beto: Tenés peritos de Parte, que te responden y estuvieron de acuerdo con los demás peritos en la autopsia. Y querés a la ONU? Q...	@Camus_Beto y eso que llevó peritos expertos. de su parte . No se confor ma , es evidente que quiere alcanzar otro objetivo.\$\$\$	2
RT @Camus_Beto: Tenés peritos de Parte, que te responden y estuvieron de acuerdo con los demás peritos en la autopsia. Y querés a la ONU? Q...	@Camus_Beto Top ten de hdp @cris maldonado hebe carlotto y la reput a madre que los parió	2
RT @cncxmendes: Me muero de la vergüenza pero acá les dejo mi cover de #Mamita - @CNCOMusic #MamitaFanVideo https://t.co/KGHF4eN1aP	@cncxmendes @CNCOMusic a penas escuche tu voz me recordaste a la usa esquivel.	2
RT @cncxmendes: Me muero de la vergüenza pero acá les dejo mi cover de #Mamita - @CNCOMusic #MamitaFanVideo https://t.co/KGHF4eN1aP	@cncxmendes @CNCOMusic Chicos miren	2
RT @SergioFVazquez: Otro clásico más y la Cato no esta en un gran mom ento deportivo pero siempre está el temple de cada jugador para poder...	@SergioFVazquez Gol tuyo !!!	2
RT @SergioFVazquez: Otro clásico más y la Cato no esta en un gran mom ento deportivo pero siempre está el temple de cada jugador para poder...	@SergioFVazquez Como cuando ganaron con 9 !	2
RT @SergioFVazquez: Otro clásico más y la Cato no esta en un gran mom ento deportivo pero siempre está el temple de cada jugador para poder...	@SergioFVazquez @CatoCampeon Y cuanto sabe Don Charly! Un abrazo mostro!! 🐼🐼🐼	2
RT @danilazarom: El director general de los Mossos, Pere Soler, se despid e por carta. Acata pues la destitución del Gobierno de la Nación....	@danilazarom @MargaPallas ¿que habria que esperar?	2
RT @patotarruella: Lo vi 48757 veces y me rei 48585884. Con y sin sonido . https://t.co/gMGVVMakbu	@patotarruella @juanitagroisman @NicoMancine me voy a morir	2
RT @robdnavarro: Este domingo a las 21 horas todo el que entre a https:// t.co/GsmiHUL21z va a poder ver el programa sin restricciones.	@robdnavarro Grande Navarro! Te necesitamos!!	2
RT @yFervazquez: Cri criminal 🇺🇵	@yFervazquez Amo esa canción	2
RT @abaftviconte: Ale Ripoll (director de SM): creo que Mica y Jorgito v an a ser la pareja de la noche👉👈	@abaftviconte dónde lo dijo??!	2
RT @Checheperez: Jorge Glas defendiendo a Odebrecht en cadena nacion al. Sabia hasta cuánto habían ofertado los otros pero ahora dice que él...	@Checheperez El país no come cuentos. Ahora el pueblo está preparado en las mejores universidades del mundo. Basta del show	2
RT @abaftviconte: Creo que Mica y Jorgito van a ser la pareja de la noc he👉👈	@abaftviconte dónde lo dijo??!	2
RT @norcorean: Si se independiza ahora Galicia se os queda tipazo. http s://t.co/5rfsGgb9fs	@norcorean @gerardotc Ponemos las Islas Canarias a la altura de Aragón	2
RT @AndresVelasq: 3 poderosas razones para convocar a gigantesca mov ilización: fraude en Bolívar, golpe en el Zulia y el hambre pareja que...	@AndresVelasq Venezuela será libre y estos asesinos, corruptos y tortura dores pasarán muchos años tras las rejas. no podrán esconderse en este p laneta.	2
RT @carolgrillito: Vuelve al Ecuador la inestabilidad, el golpismo seudo ins titucional, la troncha y los pilatos encajados en el poder temp...	@carolgrillito @MashiRafael En nuestro país no existe justicia	2
RT @carolgrillito: Vuelve al Ecuador la inestabilidad, el golpismo seudo ins titucional, la troncha y los pilatos encajados en el poder temp...	@carolgrillito @MashiRafael Fuera traidores	2
RT @carolgrillito: Vuelve al Ecuador la inestabilidad, el golpismo seudo ins titucional, la troncha y los pilatos encajados en el poder temp...	@carolgrillito @MashiRafael Los peruanos cerramos la frontera hace rato	2
RT @luciano_eltirri: Sin palabras!! 🍷 https://t.co/CiN2DmWZMO	@luciano_eltirri @cuervotinelli que cosa pasa a ser importante??	2
RT @Juliolamas1: Quiero para mi País :Q se esclarezcan los casos Maldon ado y Nisman.Q vayan presos TODOS los q robaron.Q se gobierne pen...	@Juliolamas1 Un ideal marca no soluciona los problemas.	2

Table A76: Examples of labeled data: Twitter (41)

RT @guillemochao: POR FAVOR: Ojalá nos puedan ayudar donando plaq uetas para mi padre, Guillermo Ochoa. Banco de sangre del hospital ABC de...	@guillemochao @alejandroberry Hay que tener dinero para estar allí.	2
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tard e... NAKSDFLJNJSK5K5K5 atención al siguiente tweet ♥️👀	@MicaSuarez12 Que emojis de mierda	2
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tard e... NAKSDFLJNJSK5K5K5 atención al siguiente tweet ♥️👀	@MicaSuarez12 Twittealo ya	2
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tard e... NAKSDFLJNJSK5K5K5 atención al siguiente tweet ♥️👀	@MicaSuarez12 NOS VAS A MATAR	2
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tard e... NAKSDFLJNJSK5K5K5 atención al siguiente tweet ♥️👀	@MicaSuarez12 Vas a conocer a Justin ?	2
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tard e... NAKSDFLJNJSK5K5K5 atención al siguiente tweet ♥️👀	@MicaSuarez12 Estoy filipando	2
RT @MicaSuarez12: Les voy a dar la noticia que les comenté hoy a la tard e... NAKSDFLJNJSK5K5K5 atención al siguiente tweet ♥️👀	@MicaSuarez12 estamos muriendo	2
RT @matiasmier19: Con trabajo y sacrificio, todo llega! Vamos por más, v amos @JuniorClubSAI 🇨🇴🇨🇵🇨 https://t.co/c4JpF8tqR	@matiasmier19 @JuniorClubSA Desde un comienzo dije que le tenía fe.	2
RT @patotarruella: Lo vi 48757 veces y me rei 48585884. Con y sin sonido ! https://t.co/gMGVvMAkbu	@patotarruella @MalenaStein puedes oír lo que pienso. Ñam ñam ñam ñ am ñam ñam ñam ñam ñam ñam ñam.	2
RT @LuisFlorido: Estamos en EEUU, mañana nos reuniremos con President e del TSJ en el exilio y magistrados legítimos ¡atenlos!	@LuisFlorido Lastima que ya no puedo ir. Este gobierno me secuestró o l os venezolanos todos estamos secuestrados	2
RT @LuisFlorido: Estamos en EEUU, mañana nos reuniremos con President e del TSJ en el exilio y magistrados legítimos ¡atenlos!	@LuisFlorido perdiendo el tiempo, pensando... hasta cuando...	2
RT @Maru_Abregu: En dietética Caliao cola para comprar los tés de la fa milia Maldonado. Gracias trolls, ahora tenemos una forma de ayudarlo...	@Maru_Abregu Entiendo que es bueno dar una mano.	2
RT @Maru_Abregu: En dietética Caliao cola para comprar los tés de la fa milia Maldonado. Gracias trolls, ahora tenemos una forma de ayudarlo...	@Maru_Abregu Lamentablemente no lo trae de vuelta a Santiago 🙄	2
RT @melortega1: Luego del compromiso demostrado durante la peor em ergencia del País, GFDR Media tomó la decisión de despedirme a mí y a otr o...	@melortega1 En los Estados hay más oportunidades para los bilingües .	2
RT @CampagnoliJM: Descalifica a 50 Peritos argentinos y todos los recau dos q tomó el Juez para esclarecer el hecho.cosa q no hicieron con #...	@CampagnoliJM La flia de Nisman quiere la verdad. La flia de Maldonado ... Nooo...	2
RT @AngeldebrutoOk: Muy buena noche para @flor_vigna y @Marama_A gustin https://t.co/E9JWcndyE	@AngeldebrutoOk @flor_vigna @Marama_Agustin Rock and roll al ritmo d e MJD	2
RT @basquetplus: Brillante https://t.co/sjwhOHY1dO	@basquetplus básquet o política. ? Decidanse....	2
RT @MichelPesquera: A ver... juguemos... https://t.co/3laW59dhz	@MichelPesquera Cafa, pero gustaría cafeína	2
@HeliOflores_mex: HOMBRES TRABAJANDO.Con el sudor de su frente y d e sol a sol.	@HeliOflores_mex @TIT0BE_1 No merecen mi opinión . son plebes. me a puran los dueños poderosos de los edificios mal...	2
@esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoci ón. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @maiceado o amenazado...	2
@esquivelgerardo: Ojo aquí: Santiago Nieto retira su objeción a la remoci ón. Reconoce que no hay condiciones políticas para que pudiera...	@esquivelgerardo @DeniseDresser0 Copelas o cuello	2
@mcg1465: A esta perrita sus dueños NO la quieren más!SIN PALABRAS Esta castrada, la trasladan donde alquien le de un lugarcito.Est...	@mcg1465 Gente de mierda !	2
@kevsho9: VIDEO NUEVO! Terminamos el mes de Halloween en el canal c on un sketch y estoy acompañado ...	@kevsho9 Donde hay que poner el código?	2
@BibiMonteroDR: Cómo es posible que en el extranjero sí hablen de la co rrupción en México y aquí las malditas televisoras callan! ...	@BibiMonteroDR Para llorar...	2
@GuillermoBotero: Primero suspendieron fumigación aérea y ahora quier en despenalizar siembra de 38 hectáreas de cultivos ilícitos. Dond...	@GuillermoBotero El día que elegimos a Santos.	2
@padrezurdo: @laidygomezf Estas buena y es lo unico q te salva. Deberia s vender tu cuerpo y no tu alma. Arrastrada! Bonita por fuera mas...	@padrezurdo @laidygomezf 140 caracteres sin desperdicio!! Pisoteo el he roico Táchira.	2
@eltanomdp: No olvidar a estos mercenarios oficialistas son cómplices de una desaparición forzada seguida de asesinato htt...	@eltanomdp @mononaban los monos bailan por bananas	2
@carolgrillito: Vuelve al Ecuador la inestabilidad, el golpismo seudo insti tucional, la troncha y los pilatos encajados en el poder temp...	@carolgrillito @MashiRafael Probando de su propio chocolate, salió el tir o por la culata decia mi abuelita.	2
@botedeltitanic1: SI ME AUMENTAN EL ABL NO QUIERO VER MAS EXTRA ÑEROS GRATIS EN HOSPITALES PUBLICOS. @HORACIOLARRETA	@botedeltitanic1 @horaciolarreta Basta de Salud gratuita a extranjeros.	2
@jiotegi: El presidente se ha puesto la gorra, la toga y la birome, para d ecirle a cada argentino lo que tiene que hacer (Máximo Kirchner...	@jiotegi Es un infeliz	2
@jiotegi: El presidente se ha puesto la gorra, la toga y la birome, para d ecirle a cada argentino lo que tiene que hacer (Máximo Kirchner...	@jiotegi No es el dueño. No puede ni tiene autoridad para hacer lo que se le canta	2
@pnique: Si la manifestación la convocan los que odian a los gays, a los i nmigrantes, a los rojos y a los catalanes, a lo mejor para uni...	@pnique Al revés, se los quitará.	2
@ehbildu: Tenemos que construir un muro democrático entre todos los d emócratas @AlmaldoOtegi	@ehbildu @AlmaldoOtegi Terroristas y terroristas!!!	2
@ehbildu: Tenemos que construir un muro democrático entre todos los d emócratas @AlmaldoOtegi	@ehbildu @AlmaldoOtegi @AlmaldoOtegi ¿No se te cae la cara de vergü enza?	2
@ehbildu: Tenemos que construir un muro democrático entre todos los d emócratas @AlmaldoOtegi	@ehbildu @AlmaldoOtegi Lecciones dadas por un asesino sin escrúpulos como Otegi... Todo en orden.	2
@IgorMeltvor: Sin ellos nada de esto sería posible. No lo olvidéis nunca.	@IgorMeltvor @kanallak! La risa se os convertirá el lloro Pedro Sánchez i zeta i compañía	2
@Cazatalentos: 90% de los fascistas españoles no saben que son fascistas. Los medios franquistas los han convencido de que lo suyo es es...	@Cazatalentos @sindi584 El chino de tu esquina lo sabe y le sube el pre cio a las banderitas que comprais på los balcones	2
@tortondo: Sería una vergüenza que este chiste por el que lo hizo impu tado el director de El Jueves tuviera miles de retuits.	@tortondo @orlcalvet Que pida en su defensa un drogo test	2
@OyeKarma: No le pegues a alguien donde una vez te contó que le dolía	@OyeKarma Una "integridad"... le dicen	2
@DaniMateoAgain: Un poquito de polémica de buena mañana? Venga... Opinad!	@DaniMateoAgain El pueblo tiene derecho besico y constitucional a deci dir su propio futuro, muy bien dicho Mr. Jones	2

Table A77: Examples of labeled data: Twitter (42)

@DaniMateoAgain: Un poquito de polémica de buena mañana? Venga... Opinad!	@DaniMateoAgain Es hora de que se ponga sobre la mesa un referen...	2
@DaniMateoAgain Dividir un país es un asunto muy grave con consec...	@DaniMateoAgain Lo del matrimonio no es aplicable porque estamos...	2
@VMatterfilm: No cuesta nada y les tocara las narices porque no podr...	@VMatterfilm Os están fichando a todos @policia @guardiacivil	2
@Pontifex_es: Cristo ha vencido la muerte. Él es nuestra resurrección y...	@Pontifex_es Oremos por padre.CARLOS ORACIO .de cali colombia q...	2
@AlbanoDante76: "Independentistas golpean a un chico que no quiso...	@AlbanoDante76 @gabrielrufian Seguid creando odio...seguid.	2
@eldiarios: ÚLTIMA HORA Puigdemont y los exconsellers que continu...	@eldiarios @gonzalcortizo LA JUSTICIA ESPAÑOLA VENDE JUSTICIA...	2
@numer344: Mosso amenazando al conductor de un autobús por lleva...	@numer344 @jordi_canyas Policia politica... no es la nostra.	2
@Gracias @DANIROVI...	@AMPAROSUR @DANIROVIRA preciosa y cariñosa	2
@Gracias @DANIROVI...	@AMPAROSUR @DANIROVIRA Que pena que no tenga espacio para...	2
@paolucci40: Continua ÉXODO MASIVO de Venezolanos huyendo de l...	@paolucci40 Donde estas Venezuela ?	2
@PerdigueroSIPep: Otro Mosso que no merece llevar el uniforme que...	@PerdigueroSIPep @forogc El "abogadillo" que defendia la actitud. ac...	2
@PerdigueroSIPep: Otro Mosso que no merece llevar el uniforme que...	@PerdigueroSIPep @035Maps El pobre @mossos separata está tan de...	2
@PerdigueroSIPep: Otro Mosso que no merece llevar el uniforme que...	@PerdigueroSIPep @forogc ¿Sabemos el número de placa?	2
@joan_guirado_: Una discoteca de ofrece barra libre gratis esta noche...	@joan_guirado_ @RaholaOficial @KRLS @JaumeAlonsoCuev Ahora ent...	2
@leotaxil: Hay que replantearse si tienen sentido las monarquías hoy e...	@leotaxil @eulalia23R Paco, el espadachin de PLANETA. A la los libros...	2
@leotaxil: Hay que replantearse si tienen sentido las monarquías hoy e...	@leotaxil Programas basura y noticias manipuladas!!!	2
@MariaCorinaYA: Lámelo como lo llame, Maduro anunció la bancarro...	@MariaCorinaYA Lastimosamente el juego está trancado + del 80% de...	2
@pnique: "¿Tú eres diputada?" "Para ser andaluza, eres más educada q...	@pnique Las mujeres somos iguales que los hombres. ¿Te enteras ya?	2
@jlespert: Sisí mi tesoro. Tooooda la flexibilidad laboral que quieras...	@jlespert Hace más de 16 años que vendo en Mercado Libre.	2
@jlespert: Sisí mi tesoro. Tooooda la flexibilidad laboral que quieras...	@jlespert Ojalá llegara @amazon a Argentina comparto totalmente	2
@marubimo: Guardiola, el Dalai lama pide a la comunidad internacional...	@marubimo Lo de su hermana cobrando 80mil al año en una embaja...	2
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acue...	@Radio10 La grieta es MORAL.	2
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acue...	@Radio10 Está de acuerdo que Cristóbal López pague lo que le debe...	2
@xavialerdekoa: Tocando el cielo con los dedos. Una fotografía maravil...	@xavialerdekoa Mi padre viajero constante me dice que como la luz d...	2
@geografosubjeti: En la España democrática no ha parado de haber pr...	@geografosubjeti @Tonicanto1 Cuánta injusticia	2
@PaulaVazquezTV: Y mientras sacudes tu bandera	@PaulaVazquezTV @podremos2015 Sin cumplirse la ley para la memo...	2
@PaulaVazquezTV: Y mientras sacudes tu bandera	@PaulaVazquezTV @pnique El problema es que solo ves lo que esta...	2
@PaulaVazquezTV: Y mientras sacudes tu bandera	@PaulaVazquezTV por la difusión, aunque el crédito debe llevarselo @...	2
@monlaferte: Luna, tu que lo ves...	@monlaferte Dile que se vaya a la chingada	2
@monlaferte: Luna, tu que lo ves...	@monlaferte No me hace falta sus migajas de falso amor	2
@Well086: A mi lo q + me gusta es ver como pacientes españoles q vi...	@Well086 lcs Necesito terapia para mi adicción al twitter. ...	2
@Well086: A mi lo q + me gusta es ver como pacientes españoles q vi...	@Well086 Sin ir más lejos el rey emérito y su rodilla	2
@Well086: A mi lo q + me gusta es ver como pacientes españoles q vi...	@Well086 El rey ...	2
@GINER: Nada mas que añadir, Señoría	@GINER Y muchas otras veces (casi todas) también	2
@isaacforrales: Te tienes que reir: La ultraderecha asturiana ha creado...	@isaacforrales @kj_mestre Animalicos.	2
@gabrielLucero: Valores de estética visual y estética moral	@gabrielLucero el perro en la pileta	2
@gabrielLucero: Valores de estética visual y estética moral	@gabrielLucero me ve a mi todo el día con el mate y se muere!!!!	2
@auronplay: En 2 horas es mi cumpleaños, espero que me llegue una...	@auronplay Hola soy de la empresa de wallapop	2
@Pontifex_es: Cristo ha vencido la muerte. Él es nuestra resurrección y...	@Pontifex_es Tatue texto en hebreo de la Biblia, de San Juan 11:25-26	2
@inafinogenova: ¿Alguna vez aprenderán a respetar a sus compañeras...	@inafinogenova Deja que los perros ladren	2

Table A78: Examples of labeled data: Twitter (43)

@PPTeamRafa Quiero que hagan una serie así, estaria super cool! @PPTeamRafa	@PPTeamRafa la rosa se Guadalupe o que	2
@MorenoG_Agustin: No hay mayor cobardía que la de Felipe González: 14 años en el poder y con mayorías absolutas y no hizo NADA por las v...	@MorenoG_Agustin siempre decir las verdades en su momento y no irte por las ramas	2
@MorenoG_Agustin: No hay mayor cobardía que la de Felipe González: 14 años en el poder y con mayorías absolutas y no hizo NADA por las v...	@MorenoG_Agustin @carlosbardem cuando hablamos de otros	2
@MorenoG_Agustin: No hay mayor cobardía que la de Felipe González: 14 años en el poder y con mayorías absolutas y no hizo NADA por las v...	@MorenoG_Agustin @IndEnCastellano esconderse como la rata q es en el 23F...	2
@Pontifex_es: Cristo ha vencido la muerte: El es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es Un sueño bastante simpático.	2
@PipeHenriquezO: Que importante lo que me dice este Bot. Por eso este próximo 13 de noviembre, hay que levantarse a votar, para BQTAR al...	@PipeHenriquezO Se cayó.	2
@_anapastor_: La fiscalía belga respondiendo a todas las preguntas... ¡igual que aquí!!!	@_anapastor_ en 3 idiomas y nadie cuestionándolo... años luz	2
@ForreMart: La sexta noche es el Salame de la política.	En @FAQSTV3 se habla se escucha sin gritos ni espectáculos. Construyendo	2
@SoyAmargada por favor esa carita!, me muero por Dios, gracias a toda esa gente que no les entra el corazón en el...	@SoyAmargada Bravo por la sensibilidad de la gente grande.	2
@caldostrong: Cuando tu color político no te permite ver el asiento vacío al lado de Orrego donde tu inmunda alcaldesa se puede sentar s...	@caldostrong @ejopia no es Evelyn es su papá con falda	2
@caldostrong: Cuando tu color político no te permite ver el asiento vacío al lado de Orrego donde tu inmunda alcaldesa se puede sentar s...	@caldostrong Las únicas damas estan sentadas	2
@caldostrong: Cuando tu color político no te permite ver el asiento vacío al lado de Orrego donde tu inmunda alcaldesa se puede sentar s...	@caldostrong @EduNavarroCores Me imagino la vieja facha que va de pie...	2
@DaniMateoAgain: La oposición debería preguntarse seriamente qué está haciendo mal cuando no gana ni con esto.	@DaniMateoAgain Rajoy tiene la gallina de los huevos de oro!	2
@DaniMateoAgain: La oposición debería preguntarse seriamente qué está haciendo mal cuando no gana ni con esto.	@DaniMateoAgain @Salvatwitts Por dar una idea...	2
@rcabrero75: Merece la pena que escuchéis a @OwenJones84 en la BBC sobre Catalunya. Verdades como puños...	@rcabrero75 @gabrielrufian @OwenJones84 puede decir más alto... pero más claro	2
@espuromorbo: Quién diría que con una meada empezaría todo compadre.	@espuromorbo Están Ricás	2
@AlbanoDante76: Tertuliana de @laSextaTV diciendo que la vida en la cárcel es "amable" con "siestas de dos horas". Cuanta miseria.	@AlbanoDante76 @laSextaTV Usted quiere dar miedo?	2
@AlbanoDante76: Tertuliana de @laSextaTV diciendo que la vida en la cárcel es "amable" con "siestas de dos horas". Cuanta miseria.	@AlbanoDante76 @laSextaTV Marca España	2
@SomosTigres: AAAHHH SE LES MAMÓ QUINONES, LES PINTO LA DOBLE L A LA AFICION RAYADA IAJAJA	@SomosTigres Nuevo chiste para 3 meses...-"Mi quifones se la pintó a mty yupi" Mientras Davino se las metió por el...	2
@AbelPintos: ¡BUENAS!	@AbelPintos Desde ullum San Juan!!!!	2
@lamKatieAngel: Ya estás participando para GANA E uno de estos 5 IPHONES 8 ??? ¿Qué esperas? Correleeee! http...	@lamKatieAngel Todos los días estoy poniendo por todas sus publicaciones es tan bella @lamKatieAngel	2
@lamKatieAngel: Ya estás participando para GANA E uno de estos 5 IPHONES 8 ??? ¿Qué esperas? Correleeee! http...	@lamKatieAngel saludos al oso y a cachito	2
@JeronAmethyst: No se por que no les queda claro que Army es un fandom muy poderoso, cada que nos hacen algo nos hacemos tendencia porque...	@JeronAmethyst hacemos este tipo de Hashtag	2
@JeronAmethyst: No se por que no les queda claro que Army es un fandom muy poderoso, cada que nos hacen algo nos hacemos tendencia porque...	@JeronAmethyst buscamos críticas y medias...	2
@JeronAmethyst: No se por que no les queda claro que Army es un fandom muy poderoso, cada que nos hacen algo nos hacemos tendencia porque...	@JeronAmethyst nos llaman más arenoso	2
@JeronAmethyst: No se por que no les queda claro que Army es un fandom muy poderoso, cada que nos hacen algo nos hacemos tendencia porque...	@JeronAmethyst pero ellos si con BTS...	2
@valtonyc: ¡DIFUSION! Acaba de salir la sentencia, ¡me han condenado a 3 años y 6 meses de cárcel por las canciones! ón	@valtonyc @disconnectada una puta vergüenza de país.	2
@correoderobert: Agente de la "autoridad" catalana molesto de que ondeen la bandera española dice que apunta la matrícula @forogc https...	@correoderobert @Storvice @forogc Conducir un autocar y ondear bandera s al mismo tiempo no es seguro, no?	2
@cap0: La hija de Manolo Escobar dice que su padre estaría a favor del der echo a decidir. Los comentarios a la noticia te sorprenderán...	@cap0 @tonialba el fiscal maza llama a declarar a la hija	2
@AntonioRNaranjo: Que un tipo detenido en Bruselas sea candidato en España, a una CCAA q ha hundido, por un partido que no reconoce a Es...	@AntonioRNaranjo Hay que tocar la constitución	2
@MatildaxD: Qué ser humano basura el que pide empanadas de verdura y después come de las buenas. Es el mismo que se hace el chavista de...	@MatildaxD Se hacen las hindúes y yoguis.	2
@DreambeachFest: de abono individual.Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest Hay que ir al festival más impresionante de España sin dudar!!	2
@DreambeachFest: de abono individual.Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest Mira, saquen la segunda tanda de entradas	2
@DreambeachFest: de abono individual.Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest por 5 año!	2
@DreambeachFest: de abono individual.Da a este Tweet y tweetea usando y podrás venir GRATIS a	@DreambeachFest Mi sueño es morir en la playa	2
@PaulaVazquezTV: Y mientras sacudes tu bandera ...	@PaulaVazquezTV Joder! Estás mordiendo la mano.	2
@gabrielrufian: No tenéis vergüenza.	@gabrielrufian Madre mía	2
@laufer4: 800 PROGRAMAS !!!!!!! @combatearg @canal9oficial gracias!!!	@laufer4 @combatearg @canal9oficial Hola me aceptas que se yo tu amigo si o no 1166635140	2
@villalobosbebas: no corria de ustedes, corria con ustedes	@villalobosbebas Ayyyyyyyyy llore	2
@Pontifex_es: Cristo ha vencido la muerte: El es nuestra resurrección y nuestra vida. Sean testigos de este mensaje de esperanza.	@Pontifex_es ¡ACOGES ABRAZA, CUIDA, ACOMPAÑA...!	2

Table A79: Examples of labeled data: Twitter (44)

@Sentid2016Comun: "¿Tu eres diputada? Eres muy agresiva para ser a ndaluza"Odón Elorza sin argumento ante una brillante joven política a	@Sentid2016Comun Qué desafortunado @odonelorza2011 pensé que tenía otro nivel intelectual.	2
@Sentid2016Comun: "¿Tu eres diputada? Eres muy agresiva para ser a ndaluza"Odón Elorza sin argumento ante una brillante joven política a	@Sentid2016Comun Qué desafortunado @odonelorza2011 Sera difícil es argumentar con educación?	2
@gabrielrufian: No. tenéis vergüenza.	@gabrielrufian Y el jeta no deja el Parlamento	2
@gabrielrufian: No. tenéis vergüenza.	@gabrielrufian Le dijo la sartén al cazo...	2
@gabrielrufian: No. tenéis vergüenza.	@gabrielrufian Hiciste estos carteles Rufian.?	2
@louisstaken: -mamá que hacías cuando eras adolescente?-shippeaba gays. leía fanfics. lloraba por no poder ver a mis ídolos y tenía ac...	@louisstaken @orichapame No se dan en adopción	2
@fernápaba: ¿Se acuerdan como luchó desde la vicepresidencia para mejorar los sistemas de salud y pensiones? Yo tampoco, porque nunca lo...	@fernápaba @Co_lom_bia2017 cinismo nivel máximo	2
@louisstaken: -mamá que hacías cuando eras adolescente?-shippeaba gays. leía fanfics. lloraba por no poder ver a mis ídolos y tenía ac...	@louisstaken @orichapame Después de eso le pegamos en el palo	2
@patricialopezl: Ahora????	@patricialopezl Ya ha vuelto del puente?	2
@patricialopezl: Ahora????	@patricialopezl @Manuel_Huerga Vienen elecciones, insultan inteligencia!	2
@sninobecerra: Desde el 26.09 el PRINCIPALÍSIMO problemón de ESP es la deuda pública. ¿Por qué no está en las portadas de todos los me di...	@sninobecerra Sera logico invertir en las empresas que no evaden impuestos y no gastar fondos publicos	2
@sninobecerra: Desde el 26.09 el PRINCIPALÍSIMO problemón de ESP es la deuda pública. ¿Por qué no está en las portadas de todos los me di...	@sninobecerra @lluís_llach hay que ganar elecciones	2
SEGUIRME A MI Y A @competicontrol DAR Y A LA FOTO ...	@TO88ALink @CompetiControl Soy el 999 XD.	2
SEGUIRME A MI Y A @competicontrol DAR Y A LA FOTO ...	@TO88ALink @CompetiControl Fui uno de los primeros en Instagram.	2
@gabrielrufian: No. tenéis vergüenza.	@gabrielrufian La Onu es muy clara	2
@gabrielrufian: No. tenéis vergüenza.	@gabrielrufian Que izquierda?	2
@Cristina_H_: No es que esté de moda denunciar el acoso, es que AHORA hemos generado (con esfuerzo) un contexto social en el que podemos...	@Cristina_H_ ¿Has leído el artículo en ElPais sobre la denuncia de la actriz americana, descendiente españoles	2
@RLOZINSKI: Se necesita LENALIDOMIDA en pastillas 25 mg y TRAMAL. Paciente con 9 meses requiriendo el medicamento. 0414-1004305	@RLOZINSKI @HenkelGarcia SITIOS PARA INDAGAR	2
@progadgetYT: [& amp; FOLLOWW] He llegado a 1,5M de suscriptores y para celebrarlo:SO EQ INTERNACIONAL iPhone X Participa:	@progadgetYT Quien se llevara el phone x chan chan channnn	2
@elcitizen: Se ven fuertes, se sienten invencibles, todo lo controlan. La dictadura de Maduro piensa q lo internacional no pasa mas allá...	@elcitizen @gerardo_caldera Venezuela no es Cuba	2
@alfonsopmedina: Que tome nota la Fiscalía española de lo que hacen sus homólogos en Bélgica, Francia, en toda Europa: dar ruedas de pr...	@alfonsopmedina @96_Yeyo no precisar que en tres idiomas	2
@AldoDuqueSantos: Y continúa el triste periplo de inauguraciones.Co mo si con 10 días de inauguraciones pudieran cambiar 4 años de inefic...	@AldoDuqueSantos Dígaselo a Forbes o a los beneficiarios de las obra No es necesario ser grandilocuente.	2
@La12twittera: Un clásico se cortó la LUZ en el MUDOMental	@La12twittera No tienen gas	2
@La12twittera: Un clásico se cortó la LUZ en el MUDOMental	@La12twittera aguante 80CA..!	2
@ElJuanpaZurita: Buen bello día	@ElJuanpaZurita Sigame soy nuevo en twitter	2
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracias @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Mándenme saludos	2
@dosogas: A quien le haremos la videollamada la próxima semana? (D e todas las redes sociales, los más activos)	@dosogas Podría ser no?	2
@nokeador1: "Q me separa a mi d Vertbistky y d los d los montos, q y o no mataría a nadie, y menos desarmado en 1 sótano. El...	@nokeador1 @AnitaCorts2 separa la pertenencia a la comunidad univ ersitaria.	2
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani Ya callate	2
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani Te faltó él aburrido	2
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani Ahre vente a mexico y nos vemos	2
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani Cuando nos vemos nosotros?	2
@HijoAtipico: no puedo más	@HijoAtipico La puta madre	2
@Argengchogau: AYER SÁBADO por la tarde Rio Gallegos,CFK saliendo por puerta de servicio. DESCUBIE A, vecinos le gritan chorra, ladro...	@Argengchogau @ESAlmaES Quien habla ?	2
@SomosNacional1: El impresionante golazo de Sebastián Rodríguez hoy ante Liverpool. Simplemente tremendo	@SomosNacional1 jugaron con el instituto del desorden	2
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracias @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Son mis Youtubers favoritos	2
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracias @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ , Saludos a los dos	2
@Juandedios_P: Estoy amando mucho este proyecto de JUKILOP gracias @KimberlyLoaiza_ por la iniciativa	@Juandedios_P @KimberlyLoaiza_ Salúdame plis	2
@BeatrizTalegon: La población Catalana tiene una cultura democrática muy por encima de la media española. Por eso aquí no se entiende ca...	@BeatrizTalegon ¿La cultura política depende de dónde estes empadronado?	2
@BeatrizTalegon: La población Catalana tiene una cultura democrática muy por encima de la media española. Por eso aquí no se entiende ca...	@BeatrizTalegon Enseñar a los pobres españoles de mierda	2

Table A80: Examples of labeled data: Twitter (45)

@soycrismorena: Gracias @elotroladock, la risa sanadora provocada por excelentes y queridos actores en una comedia que no se pueden perd...	@soycrismorena @ElOtroLadoOk Haz algo nuevo	2
@soycrismorena: Gracias @elotroladock, la risa sanadora provocada por excelentes y queridos actores en una comedia que no se pueden perd...	@soycrismorena @ElOtroLadoOk Sería un éxito	2
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer...	@missencuestas2 tengo 3 perros y dos gatos	2
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer...	@missencuestas2 tengo 5 rescatadas pero una se orinaba del miedo que...	2
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer...	@missencuestas2 no seguir de largo	2
@missencuestas2: En Diciembre del año pasado me lo encontré casi muer...	@missencuestas2 Tengo 5	2
@FSerranoCastro: En este país es más peligroso rebelarse contra la dictad...	@FSerranoCastro los lobbys protegidos e impulsados desde los órganos...	2
@FSerranoCastro: En este país es más peligroso rebelarse contra la dictad...	@FSerranoCastro Ejemplos tenemos....	2
@Chileokulto: Jajajajaja... de colección, así desenmascaró periodista a @s...	@Chileokulto @sebastianpinera Se graduó en Harvard	2
@BeatrizGH_17: Holi mis pajaritos Hoy será un gran día, empezamos por...	@BeatrizGH_17 Con ganas de ver en el debate	2
@Miotroyo2parte: ¿Cómo se financiarán tras la separación de España? R...	@Miotroyo2parte @FaltaVendimia @martarovira Creía que "La chica del...	2
@DreambeachFest: de abono individual.Da a este Tweet y tweeketa usand...	@DreambeachFest poder repetir mañanas ...	2
@fedevigevani: Digo esto porque no me gustan los domingos	@fedevigevani Odio los domingos. Son mas aburridos.	2
@gabrielrosmega: Anoche murió un amigo. 36 años. Le dió un infarto y u...	@gabrielrosmega A veces laboralmente nos creemos imprescindibles	2
@JOSEMANUELSTOTO1: Franquistas en España no quedan, apuestan mucho...	@JOSEMANUELSTOTO1 @inakilabella El Perez Royo no sabe que hacer pa...	2
@ElisendaPerez: También decías que eras el líder para acabar con el gobi...	@ElisendaPerez @gabrielrufian Sin palabras verdades como puños	2
@Coordinadora25S: Madrid está con el pueblo catalán. Desde la Puerta d...	@Coordinadora25S Muchas gracias por vuestro apoyo!!!	2
@Coordinadora25S: Madrid está con el pueblo catalán. Desde la Puerta d...	@Coordinadora25S Que mal cuentas no? Será algunos	2
@auronplay: Por fin me han puesto los 280 caracteres soy muy feliz en es...	@auronplay esto ya es racismo	2
@Camila_Cabello: Te amo.....	@Camila_Cabello Sigue por favor	2
@Camila_Cabello: Te amo.....	@Camila_Cabello hola me llamo karla y soy tu tocaya	2
@JuanEnriquez5W: Ayer hinchas de la U asistieron a los hinchas de Wand...	@JuanEnriquez5W Rivaletas, NO enemigos!	2
@TomiOlava: Yo no quiero ser un opositor irracional pero realmente no p...	@TomiOlava Pensó que era un scrum.	2
@_anapastor_: No me canso de verlo...	@_anapastor_ Otra cosa no será pero sincero...	2
@_anapastor_: No me canso de verlo...	@_anapastor_ Hijo de la gran puta	2
@Anitta: Hola queridos! Envíenme sus preguntas! Te imaginas si respond...	@Anitta Tienes ganas de hacer featuring con cual cantante?	2
@brozoxmiswebs: Dice el refrán: cuando el río suena, es que agua lleva, a...	@brozoxmiswebs @calbert57 Que pasara con los implicados? Van a recib...	2
@1MaduroChavista: Un Majunche Amargado Queriendo insultar a un CH...	@1MaduroChavista @Deci3003 @Zikdal @dblancong o tratando de meter...	2
@camilaacarasso: Que deprimente ponerse la bikini maaaamita	@camilaacarasso Seguis siendo mi amiga despues de esta fama?	2
@RoMorales_: Foto del día, la mejor.	@RoMorales_ ¿Quien lo condena?	2
@LuisCarlos: La RAE reconoce el término venezolano "mamagüevo" y hay...	@LuisCarlos Simplemente eres único, I Love You.	2
@yanina_rubia: CASO LA MANADA:El juez rechaza como prueba los What...	@yanina_rubia @maneln51 Volvemos al tiempo de franco, donde el mac...	2
@Rubius: Doblaje Español confirmed	@Rubius POR ESTÁS COSAS TE QUIERO	2
@Pedro_Arancon: Es una pena que los cientos de banderas desplegadas...	@Pedro_Arancon Ayer el Sr. Granados estuvo preparando su defensa, al p...	2
@Pedro_Arancon: Es una pena que los cientos de banderas desplegadas...	@Pedro_Arancon Hombre xd que cosas dice usted Don Pedro.	2
@Pedro_Arancon: Es una pena que los cientos de banderas desplegadas...	@Pedro_Arancon un Patriota y ellos són putos representantes de mierda...	2
@RuidoEnLaRed: El Auditorio Metropolitano de Tecamac, que se inaugur...	@RuidoEnLaRed despues de queja de Bulling abusan del pobresito	2
@pccomponentes: SOMOS 3º A NIVEL NACIONAL EN TENDENCIAS! Com...	@pccomponentes @OMENbyHP toca cambiar el equipo jehehe	2
@pccomponentes: SOMOS 3º A NIVEL NACIONAL EN TENDENCIAS! Com...	@pccomponentes @OMENbyHP hoy es el día para vivir un futuro promet...	2
@pccomponentes: SOMOS 3º A NIVEL NACIONAL EN TENDENCIAS! Com...	@pccomponentes @OMENbyHP estaría bien renovar el ordenador	2

Table A81: Examples of labeled data: Twitter (46)

@pccomponentes: SOMOS 3º A NIVEL NACIONAL EN TENDENCIAS! C	@pccomponentes @OMENbyHP o como afilarse los dientes antes de l	2
omenta con y sigue a PcComponentes y @OMENbyHP y mu...	as propias rebajas	2
@RubiUS: Doblaje Español confirmed	@RubiUS Me vas a comer el coño con pan bimbo	2
@RubiUS: Como joder la saga entera de Pokemon en 30 segundos	@RubiUS En realidad 0:29 Pero...SI QUE LA JODIERON!!	2
@labeasanchez: Y pensar que todavía hay gente que cree que la única	@labeasanchez Educación Inversión, por lo tanto se debe pagar.	2
forma de comprometerse con algo es pagando...		
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana... :)	@ecosdelbalon @SpheraSports @XabiAlonso Grande Gasol	2
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana... :)	@ecosdelbalon @XabiAlonso Como me alegra que vuestro producto si	2
	ga creciendo y triunfe gente que cuida este deporte...	
@Schuma78: ¡¡BRUTAL!! Susanna Griso rebate los argumentos a líder i	@Schuma78 @FrayJosepho El señor es un maleducado	2
ndependentista y este la manda a casa a limpiar la ropa y hacer la...		
@MariaBecerra22: Te miro y pienso... que suerte que tengo de tenerte,	@MariaBecerra22 Nadie me ama	2
que suerte de haberte encontrado de casualidad, que suerte de que...		
@PeroniaRepublic: Che, quien es Sietecase? Esto:	@PeroniaRepublic Un mal tipo...	2
@PeroniaRepublic: Che, quien es Sietecase? Esto:	@PeroniaRepublic @TatideVCP Este tipo se merece perder su trabajo y	2
	no encontrar otro.	
@Well086: A día de hoy el estado español lleva gastados 15 Millones d	@Well086 @anaraljaan ¿Que función o sentido tienen ahora?	2
e euros en los Píolines de CAT. 15 millones! Cuántas escuelas son e...		
@PatoBullrich: Adulteró el tanque de nafta para esconder 30 kilos de c	@PatoBullrich @gendarmaria ¿quién ahogó a Santiago Maldonado?	2
ocaina. Los llevaba de Salta a Santa Fé. Lo detuvimos e incautamos...		
@PatoBullrich: Adulteró el tanque de nafta para esconder 30 kilos de c	@PatoBullrich @gendarmaria Otra mas y van . siga metiendole las man	2
ocaina. Los llevaba de Salta a Santa Fé. Lo detuvimos e incautamos...	os donde mas les duele .	
@DrEnriqueParis: Los precios de los medicamentos en Chile son mayor	@DrEnriqueParis @sebastianpinera Es harto papeleo para importar un f	2
es de lo que realmente corresponde. Son más caros que en Argentina.	ármaco .	
E...		
@ChileBicampeon: increíble como se descontrola el delincuente Piñera	@ChileBicampeon Axel Kaiser se los come a los dos con zapatos	2
cundo alguien le demuestra que está mintiendo, alz...		
@McCArni: Estamos todos de acuerdo en que Soda Stereo es la mejor	@McCArni En que influye eso en la vida de la gente ?	2
banda de rock en español de la historia de la humanidad, verdad?		
@srtabebi: No me ha bajao.	@srtabebi @albasdf La has metido ya?	2
@srtabebi: No me ha bajao.	@srtabebi El crush tiene novia	2
@srtabebi: No me ha bajao.	@srtabebi Tengo que hablar contigo	2
@srtabebi: No me ha bajao.	@srtabebi Te pasa algo conmigo?	2
@XarnegoSedicios: Junker amb una taja inhumana. Ojo al final del vid	@XarnegoSedicios en su estado natural	2
eo quan arriba Victor Orbán (PM d'Hongria)		
@2010MisterChip: Le veis clasificándose para su sexta Copa del Mundo	@2010MisterChip Hora de despertar de la siesta	2
en una tanda de penaltis? Yo me lo he imaginado y ha sido precioso...		
@2010MisterChip: Le veis clasificándose para su sexta Copa del Mundo	@2010MisterChip ¿Qué le pasa a Italia ?	2
en una tanda de penaltis? Yo me lo he imaginado y ha sido precioso...		
@2010MisterChip: Le veis clasificándose para su sexta Copa del Mundo	@2010MisterChip Sería un sufrimiento con felicidad	2
en una tanda de penaltis? Yo me lo he imaginado y ha sido precioso...		
@ErnestoEkaizer: Soy, según este diario, "otro argentino en nuestra vid	@ErnestoEkaizer No hay mayor desprecio que no hacer aprecio	2
a política". Este es el estilo editorial nacionalista es...		
@nelsonbocaranda: Rafael Ramírez montó en los pasillos frente al Cons	@nelsonbocaranda En el país la gente muriendo por falta de medicam	2
ejo Seguridad ONU exposición acusando a la oposición de todos los m	entos.	
@nelsonbocaranda: Rafael Ramírez montó en los pasillos frente al Cons	@nelsonbocaranda @Trafico Cuánto costó eso?	2
ejo Seguridad ONU exposición acusando a la oposición de todos los m		
@nelsonbocaranda: Rafael Ramírez montó en los pasillos frente al Cons	@nelsonbocaranda @Trafico Podría utilizarse para medicamentos.	2
ejo Seguridad ONU exposición acusando a la oposición de todos los m		
@cristobalsoria: Descubre las 7 diferencias...	@cristobalsoria Si es que aunque sea del Madrid me tengo que reir co	2
	n el cabr n este . Que arte	
@cristina_pardo: Rajoy dice, en alusión a los políticos independentistas,	@cristina_pardo @gabrielrufan Pienso que solo quedarían algunas muj	2
que deberían estar inhabilitados por mentir. Como ponga ahí el...	eres políticas. Y no sé si muchas.	
@guillier: Prepárense todos, que si la derecha gana será la persecución	@guillier Terror terror terror terror terror, no tengas miedo.	2
más brutal, porque el diálogo que han prometido es falso, no van...		
@guillier: Prepárense todos, que si la derecha gana será la persecución	@guillier viene el fin del mundo	2
más brutal, porque el diálogo que han prometido es falso, no van...		
@matiasdelrio: Este es el país que hemos construido y del que algunos	@matiasdelrio En poco tiempo mas llegaremos a los 120 como calcula	2
sienten vergüenza.	n las AFP.	
@pccomponentes @OMENbyHP xq nunca me toca nada	@pccomponentes @OMENbyHP La segunda es la buena	2
@pccomponentes @OMENbyHP xq nunca me toca nada	@pccomponentes @OMENbyHP Menudo sorteo muchas gracias y su	2
	erte	
@vox_es @jusapol Muchas gracias x vuestro apoyo	@vox_es @jusapol A llenar Madrid	2
@rafaellacava10: Buenos días a todos. El carro de Drácula no duerme y	@rafaellacava10 @gabrielprato Los galpones llenos de comida?	2
esta madrugada agarro a estos pajaritos q usaban una ambulancia pa...		
@rafaellacava10: Buenos días a todos. El carro de Drácula no duerme y	@rafaellacava10 @gabrielprato Los repartieron? No? Se pudrió esa co	2
esta madrugada agarro a estos pajaritos q usaban una ambulancia pa...	mida?	
@cristobalsoria: Descubre las 7 diferencias...	@cristobalsoria Nunca vi una persona más irrespetuosa que tú !	2
@Pablo_Iglesias: Pluracontradicción!! En Democracia decide la mayoría	@Pablo_Iglesias Arrepentidos los quiere La República.	2
? O una minoría? Ahh no... que tú hablas de dict...		

Table A82: Examples of labeled data: Twitter (47)

@MoruRossi: Pileta de noche, terraza de noche, balcón de noche, caminata de noche. Cualquier cosa si se hace de noche garpa mil veces m...	@MoruRossi Ya superaste a tu hermana hasta en la fama	2
@Jlespert: Nuestros problemas urgentes de corto plazo son un déficit fiscal y externo insostenibles e inflación altísima que no baja, pe...	@Jlespert Sobran 15 millones de negros parásitos.	2
@BeltraminCarlos: Si Tenenbaum y Sietecase se querían despegar del Kirchnerismo, Luisito los dejó más pegados que ventosa...	@BeltraminCarlos @aries2011956 Para que me salude este personaje prefiero no ganar nada	2
@pccomponentes: SO EAZO! Comenta con y sigue a PcComponentes y @OMENbyHP y muy importante rellena tus dat...	@pccomponentes @OMENbyHP Estoy cruzando los dedos por encima de mis posibilidades	2
@jg_dominguez: En La Vanguardia dicen ahora que la culpa de todo fue de... Rufán. Al final, y pase lo que pase, el malo siempre es un c...	@jg_dominguez Qué bonito lo de las pantallas de los móviles.	2
@pccomponentes: SO EAZO! Comenta con y sigue a PcComponentes y @OMENbyHP y muy importante rellena tus dat...	@pccomponentes @OMENbyHP Quien me regala esto?	2
@pccomponentes: SO EAZO! Comenta con y sigue a PcComponentes y @OMENbyHP y muy importante rellena tus dat...	@pccomponentes @OMENbyHP Lo tenéis en azul?	2
@DanielRetamozo: Jugó 5 Mundiales. Ganó 19 títulos locales y 3 internacionales. Es millonario. Es Campeón del Mundo. Tiene 39 años y...	@DanielRetamozo No es Pasión ... es Fútbol!	2
@perezreverte: Sobre equiparación salarial PN y GC con otros cuerpos, como en tantas otras cosas. Mariano Rajoy ya prometió (y mentó) e...	@perezreverte desde el primer día que le pedí un retuit a JUSAPOL ha estado ahí.	2
@guruchuirer: ¡Insuadito! Ningún presidente del mundo se queja de lo que se le queja Peña Nieto. En todos los países hay redes sociales, soc...	@guruchuirer @JohnLo36732656 Todavía se enoja el inepto.	2
@ViviArrow: La Manada: Angel Boza, Jesús Escudero, José Ángel Prenda, Alfonso Jesús Cabezuolo y Antonio Manuel Escudero. Un guardia civil...	@ViviArrow @XSlaimartin Cuando se trata de cerdos se llama "piara".....	2
@daughterlarryr: Gianluigi Buffon se va retirar en el mundial del 2018 e Italia va quedar campeona, guarden este tweet.	@daughterlarryr Un cordial gafe a Pablo Sobrado	2
@Davanana: VAYANSE A LA VERGA	@Davanana Eres hermosa!	2
@Davanana: VAYANSE A LA VERGA	@Davanana quedado dañadas por eso, se los dicen una vez y su ego se infla descontroladamente	2
@Rubius: Doblaje Español confirmed	@Rubius NO SE PUEDE CON VOS UNA CHICA INTENTANDO ESTUDIAR Y S	2
@ELBRAYAN: Nunca se preocupen si los critican. Al final la gente suele criticar absolutamente todo. Hagan lo que quieran, estudien lo...	@ELBRAYAN_ Iba muy bien hasta que dijo barbie girl	2
@chochos: Andele para que se le quite...	@chochos La mejor foto novela ever	2
@FonsiLoaiza: Hoy los chavales de Alsasua cumplen un año en la cárcel, sin juicio, por una trifulca de bar. Mientras a ellos los dejan q...	@FonsiLoaiza @francis4973455 son unos cobardes y unos cerdos radicales	2
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador. Baltazar Ushca, de 76 años, es la única persona que aún recolecta...	@elcomerciocom Verdadero DOCTORIS AL HOMBRE DE TRABAJO MIREN C	2
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador. Baltazar Ushca, de 76 años, es la única persona que aún recolecta...	@elcomerciocom DOCTORIS DE ROBO	2
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador. Baltazar Ushca, de 76 años, es la única persona que aún recolecta...	@elcomerciocom En doctorado en que? Como es eso posible.no...	2
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador. Baltazar Ushca, de 76 años, es la única persona que aún recolecta...	@elcomerciocom Que gustó saber eso.	2
@elcomerciocom: Un reconocimiento a un oficio emblemático en Ecuador. Baltazar Ushca, de 76 años, es la única persona que aún recolecta...	@elcomerciocom deberían darle eso	2
@SDHArgentina: Dale y ayudanos a seguir difundiendo derechos para hacer entre todos una sociedad más inclusiva e igualitaria. ht...	@SDHArgentina @JProArg No respetar a los que tienen valores cristianos s	2
@SDHArgentina: Dale y ayudanos a seguir difundiendo derechos para hacer entre todos una sociedad más inclusiva e igualitaria. ht...	@SDHArgentina @JProArg la ley no se basa en valores éticos o morales	2
@luisvcenteleon: Ligar por un default de Venezuela y pensar que eso sacará al gobierno es una pendejada... al cuadrado.	@luisvcenteleon dialogo hasta en 2040 ?	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal LA UNIÓN DE LOS PERONISTAS PUEDE C	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @ALFREDOLEUCO @FernandezAnibal Quédate tranquilo que dentro de poco	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @ALFREDOLEUCO @FernandezAnibal Seguramente lo vas a ver todos los días...	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @omarlavieri @FernandezAnibal Lo fue a visitar o a conocer las instalaciones para cuando le toque entrar esposado a él??	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal Tenes que devolver la guita	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal Sólo fue a ver cómo será su próximo domicilio	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal No puedo creer que delinquentes desde la cárcel, usen tweeter. OJO	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @AngeldebrittoOk @FernandezAnibal Está preso, porque maneja redes sociales ?	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal Parece un mensaje de los deudos...	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @FernandezAnibal Actúan los verdaderos amigos	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @omarlavieri @FernandezAnibal Debe haber ido para conocer	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @omarlavieri @FernandezAnibal Tiene acceso a internet, PC celu en la cárcel?!	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @mis2centavos Dice @FernandezAnibal Ponen aire o no va más!	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @AngeldebrittoOk @FernandezAnibal son peores que las minas con el puteri...	2
RT @Schuma78: ERC en público: "Esto no va de independencia, va de demostración" ERC en privado: "Por la independencia nos pasamos la democracia..."	@Schuma78 La "señora" debía ser cesada inmediatamente	2

Table A83: Examples of labeled data: Twitter (48)

@Schuma78 Y se puede saber quién es esta señora demócrata, que muestra esa voluntad de ayudar... Sin decirlo porq... https://t.co/sf5nVzeMCF	@Schuma78 siguen...suma y sigue...que asco.!	2
RT @JCTapiaLMB: Amenazar a los diputados para que voten a favor de la propuesta del Ejecutivo en las designaciones a la CSJ, violando la in...	@JCTapiaLMB Sencillo Don Juan Carlos, el presidente Varela se quiere "blindar" para evitar posteriormente ser inves... https://t.co/LBkez273K3	2
RT @JuanMorenoRe: Hemos recusado al nuevo magistrado designado para juzgar la caja B del Partido Popular. Su carrera está estrechamente lig...	@JuanMorenoRe Harán lo que les salga de las narices, pero que sepan que no nos tragamos sus trapicheos.	2
RT @JuanMorenoRe: Hemos recusado al nuevo magistrado designado para juzgar la caja B del Partido Popular. Su carrera está estrechamente lig...	@JuanMorenoRe Imparcialidad e independencia judicial?	2
RT @JuanMorenoRe: Hemos recusado al nuevo magistrado designado para juzgar la caja B del Partido Popular. Su carrera está estrechamente lig...	@JuanMorenoRe @Manu_Abu_Carlos Si ea dudai\$ por dudAr...desconfiados #QuerellaBarcen	2
RT @JuanMorenoRe: Hemos recusado al nuevo magistrado designado para juzgar la caja B del Partido Popular. Su carrera está estrechamente lig...	@JuanMorenoRe @agarzon Fuera corruptos	2
RT @balancelasauce: Mi hermano y yo somos la primera generación de catalanes de una familia. 100% gaditana. Todos votaremos independencia. L...	@balancelasauce @InesArrimadas lastima de padres luchadores	2
RT @ChairoMexicano: La bruja anulista @DeniseDresserG, que se sumó al frente PAN-PRD-MC, dice en Aristegui en Vivo que AMLO es "protopopuli...	@ChairoMexicano @DeniseDresserG @Irma_Sandoval @JohnMackerman le paga la CIA para dividir al pueblo.	2
RT @ChairoMexicano: La bruja anulista @DeniseDresserG, que se sumó al frente PAN-PRD-MC, dice en Aristegui en Vivo que AMLO es "protopopuli...	@ChairoMexicano @DeniseDresserG @Irma_Sandoval @JohnMackerman Se menciona que es agente de la CIA	2
RT @ChairoMexicano: La bruja anulista @DeniseDresserG, que se sumó al frente PAN-PRD-MC, dice en Aristegui en Vivo que AMLO es "protopopuli...	@ChairoMexicano @DeniseDresserG @Irma_Sandoval @JohnMackerman Pinche chaqueta de mierda	2
RT @CCivicaCatalana: Atacan el despacho de unos abogados por oponerse a l proceso. "No somos partidarios de la independencia y en Cataluña só...	@CCivicaCatalana Triste coincidencia https://t.co/NFHzCWsiy	2
RT @Protectoresazul: "San Miguel Arcángel guía con tu espada de Justicia a todo hombre y mujer que en Nombre de DIOS Todopoderoso, Nuestro Se...	@Protectoresazul @franacfra Sigán creyendo esas pendejadas. sigan que con eso van a tumbar a la dictadura. bolsai	2
RT @Schuma78: ERC en público: "Esto no va de independencia, va de democracia" ERC en privado: "Por la independencia nos pasamos la democrac...	@Schuma78 A la Fiscalía y a la Junta Central Electoral...	2
RT @Pablo_Iglesias_: "A muchos les molestarán estas palabras de Juliana de hoy. Las verdades, a veces, son amargas y difíciles de tragar Q...	@Pablo_Iglesias_ Tais tots junts en la palabreria vacua y nazi/comunistoide. Miento, falta algún ayatollah	2
RT @Vitalisioo: RT si la ves gris y celesteMG si la ves rosa y blanca https://t.co/8HFeAsfWgy	@Vitalisioo Fin de la discusión xD https://t.co/lwJ8a2EC55	2
RT @Wewerwerwo: Más que karma es porque ellos se parten el lomo contra Brazil, Italia, Colombia, Argentina, etc. Y Mex... https://t.co/UboAaUmc8z	@Wewerwerwo Que hicieron ellos?	2
RT @SilvestreFDC: Al final me quedo con las mujeres q les gustó, no con las q me gustan! Triste terminar de hacer el amor y no tener de que...	@SilvestreFDC qué es eso?	2
RT @yFervazquez: Cni criminal [7]	@yFervazquez Hoy se sale????	2
RT @mauriciomacri: No hay discusión. Messi deslumbra siempre. Aplausos para un jugador de otro mundo. [8]	@mauriciomacri @slgarelli_cnn Presidente compraron el partido dejense de joder	2
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/8HFeAsfWgy	@pccomponentes @elchapuzas no quiero morir por favor	2
RT @pccomponentes: Se acaba! RT+Follow y comenta con #AllwaysPcComponentes Llévatelo con PcComponentes y @elchapuzas Bases sorteo en https://t.co/8HFeAsfWgy	@pccomponentes @elchapuzas Quiero ser feliz	2
RT @Lhuchin: Chileno, no te olvides que el CSM que sacó a Bielsa fue Sebastián Piñera.	@Lhuchin @lamesadelcafe 0 aporte el comentario	2
RT @Oficialmorris: Un PRI que se multiplicó y tuvo muchos mini PRIs https://t.co/9r8aFbKtbp	@Oficialmorris hay hijastros que son rebeldes	2
RT @CARLOSGLUERRERO: #WorldCup2018 #WorldCup #Russia2018 Robben, vamos por la revancha en Rusia!! ... aaah... uy, bueno... https://t.co/ORL5ex0qoA	@CARLOSGLUERRERO Contra los que México no puede en el proceso de sorio	2
RT @NoblexAr: Aclaremos desde ahora: La #PromoNoblex no era válida para Chile [9] #Rusia2018 https://t.co/ZMKxC4La78	@NoblexAr es más barato comprar un Tv...	2
RT @GianThurber: Los ticos son los mejores vecinos que la vida nos pudo dar, celebraron y alentaron a la sele como un panameño más.	@GianThurber jugaron a media máquina #ticos #puravida	2
RT @La12twittera: Una DEDICATORIA de los jugadores de @argentina para la PRENSA argentina https://t.co/0g372KQWf	@La12twittera @MiliCabj_5 @Argentina Cuanto le pagaron a Ecuador?	2
RT @ElMalaguero: POR FAVOR, LID. https://t.co/FHCUPjhASR	@ElMalaguero Como me puedo bajar la genialidad?	2
RT @ZassCometoro: El diario @elmundoes nos regala un artículo de la más depurada técnica de editorial cuñado y fuentes sin contrastar.Luego...	@ZassCometoro @elmundoes @emergenciescat El Mundo nunca ha pasado de papel higiénico impreso.	2
RT @TomasCivetta: Que partido te perdiste Maldonado https://t.co/dRzxxSosl	@TomasCivetta qué falta de respeto y empatía joder con la desaparición de una persona, cerebros de casa.	2
RT @JoseLChilavert_: @joaquinfalco La de tu esposa mejor, es mas rica	@JoseLChilavert_ @joaquinfalco de donde salió	2
RT @AlejandroSanz: De ustedes depende para qué utilizar sus manos... acariciar, abrazar o dar palmas [10] https://t.co/kPwSQ2NBf	@AlejandroSanz En Costa Rica hay mucho desastre y damnificados por las inundaciones	2
RT @RuidoEnLaRed: Pueden mostrar la factura y la declaración respectiva? Quié n les dio acceso?	@RuidoEnLaRed En ese carro está el dinero del pueblo.	2
RT @lavozdegalicia: [11] El Valle de Arán anuncia un pleno el lunes para independizarse de Cataluña https://t.co/xFd3ndt8 https://t.co/wUTvD...	@lavozdegalicia voy a baqueira el puente de diciembre	2
RT @abarceloh25: Los senadores del PP jalean a las medidas del 155 mientras Rajoy las enumera. [12]	@abarceloh25 @eulalia_fe No odiamos a España, tan solo pedimos que se nos respete...	2
RT @Nomaskmail: Echarrí no se fue del país.Brancatelli tampoco.Víctor Hugo menos.Hebe no se suicidó.Promesas de campaña incumplidas.	@Nomaskmail Querrán un incentivo?	2
RT @Nomaskmail: Echarrí no se fue del país.Brancatelli tampoco.Víctor Hugo menos.Hebe no se suicidó.Promesas de campaña incumplidas.	@Nomaskmail @BrunoRubnRuizD1 son amigas y alumnas de cristina	2
RT @cap0: Rajoy: "Violar las leyes, y por tanto la democracia, tiene consecuencias". ¡AY, QUE VA A DIMITIR!	@cap0 por un momento me he hecho ilusiones...	2
RT @Blue_Leader_: #DuiARVRECONOCIMIENTO INTERNACIONAL CATALUÑA https://t.co/WYQ851rkSW	@Blue_Leader_ Ahora dónde jugará Piqué?	2
RT @Blue_Leader_: #DuiARVRECONOCIMIENTO INTERNACIONAL CATALUÑA https://t.co/WYQ851rkSW	@Blue_Leader_ Quieren independencia, entonces que el Barcelona no juegue más en España.	2

Table A84: Examples of labeled data: Twitter (49)

RT @ilseCeroUno01: Llorando, estoy llorando os lo juro. https://t.co/R36NNiWx8J	@ilseCeroUno01 @El_Formalito Cuando tus ideales te impiden ver la realidad. Y se puede aplicar a casi todos los aspectos de la vida	2
@resistenciacaata Claro que sí!!!! Que oigan a esa gente que no está de acuerdo! 🇨🇺🇨🇺🇨🇺🇨🇺	@resistenciacaata Puedo pedir un temita?	2
@resistenciacaata Claro que sí!!!! Que oigan a esa gente que no está de acuerdo! 🇨🇺🇨🇺🇨🇺🇨🇺	@resistenciacaata En manresa va sonar bien fuerte	2
RT @ensanro: La Declaración Unilateral Independencia es un error político que hará retroceder a #Catalunya y fortalecerá a la derecha española...	@ensanro @Caninorator España es un Estado fallido.	2
RT @Elaguiguij: A ver lo que hace mañana este gilipollas sin DNI. https://t.co/st1LNM7nZ6	@Elaguiguij_ ¡Adeu, 11€!	2
RT @edufeiok: El parlamento se limpia un poco https://t.co/vZ2sfD8QBd	@edufeiok @arielmnt1 Quien es el suplente de Devido?	2
RT @1Eme: Sitios donde se reconoce la República Independiente de Cataluña #DuiARV #IndependenciaCatalunya https://t.co/P7JaZND5LA	@1Eme_ Quien rie ultimo rie mejor	2
RT @Cumparini: Me topé a Kast saliendo de Santiago por el finde largo https://t.co/kNL7depNON	@Cumparini Pucha que son graves, era una broma	2
RT @luis_sisepuede: Enhorabuena, ya no tendréis que elegir entre la paleta azul, la naranja y la roja para votar a la derecha #L6ARVCat ht...	@luis_sisepuede A la derecha se le tiene que poner las cosas fáciles, no sea que se lesionen pensando.	2
RT @kevsho9: VIDEO NUEVO! 📺 Terminamos el mes de Halloween en el canal con un sketch y estoy acompañado 📺 https://t.co/iRLPIMtdDf #Hallowee...	@kevsho9 Donde hay que poner el código?	2
RT @carnecrutaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrutaradio @emparmoliner Creo que a las próximas, el Borbón no se presenta.	2
RT @carnecrutaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrutaradio navidad está cerca	2
RT @Helioflores_mex: HOMBRES TRABAJANDO. Con el sudor de su frente y de sol a sol. https://t.co/a6hiilmZv2	@Helioflores_mex @TITOBE_1 No merecen mi opinión, son plebes.	2
RT @carnecrutaradio: Ya que hemos abierto el melón de la República, ¿para cuándo un referéndum para proclamarla en España? #TerceraRepública	@carnecrutaradio @emparmoliner Está al caer.	2
RT @matiaslongoni: En septiembre se exportaron 20 mil toneladas de carne vacuna (31 mil toneladas peso res). Es el volumen mensual más alto...	@matiaslongoni Cambio de tema.	2
RT @renfrancis: Ya casi es Noviembre y yo sigo sin ti.	@renfrancis Vuelvan a MONTERREY	2
RT @CUBONoticiasMX: 📺 Mira @alfredodelmazo así "tratan" a los reos del Penal Neza Bordo. ¿Algo que decir? 📺 https://t.co/uspCmVli2t	@CUBONoticiasMX @alfredodelmazo Firma la petición	2
@JavierMMoure @jinsua49 Con esa vocisita de nenita, hacé shhhhh y no jodas más. https://t.co/pRoROJkIx5	@JavierMMoure @ArielJob Leo todo lo que ponen de Sergio Maldonado y no veo críticas. Son simples apreciaciones!!!	2
RT @NicolasMDk: Cayó un aguaviva en la ventana del edificio. https://t.co/07OLB4yuWE	@NicolasMDk @PiNKeAdA Eso cuenta como tener sexo?	2
RT @NicolasMDk: Cayó un aguaviva en la ventana del edificio. https://t.co/07OLB4yuWE	@NicolasMDk @amitriaviesas no respetan nada	2
RT @LaurenJauregui: Lo que quieras si me dices Por Favor https://t.co/4eMm4egUw	@LaurenJauregui @arianasharmoni vuelves a argentina?	2
@ziberia: Casi me echan del bar de la risa	@ziberia Ahora voy a dominar el mundo !!!	2
@DaniMateoAgain Dividir un país es un asunto muy grave con consecuencias, no debe ser tratado tan "emocionalmente" como hace él.	@DaniMateoAgain Hablan de fronteras y UE hay países que tienen millones habitantes que Málaga. Por ej. Luxemburgo.....	2
@ju1_: Tras acabar una marcha españolista en Mataró. "Di viva España o te matamos". El agredido, hospitalizado. Difunde	@ju1_ @soler_toni Donde estaba la policía?	2
@elena2760 Que tiempos aquellos	@elena2760 @adeladessire ¿Cómo se dejaron quitar eso?	2
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo o que debe abonarlo para verlo?	@Radio10 Cual es el motivo por el cual deba ser gratis?...el cine es gratis?...el teatro es gratis?	2
@Radio10: ENCUESTA: Hoy es el 1º Superclásico pago. ¿Está de acuerdo o que debe abonarlo para verlo?	@Radio10 No porque el fútbol es un derecho...	2
@xaverialdekoa Tocando el cielo con los dedos. Una fotografía maravillosa	@xaverialdekoa Mi padre viajero constante me dice que como la luz de África en ningún sitio.	2
@Samsung_Peru: Porque creemos en ti	@Samsung_Peru Ha sido siempre el capitán de selección peruano;	2
@estresfx_ dismenorrea, endometriosis, embarazo ectópico, ovario poliquístico... ¿sigo?	@estresfx_ @Staptra No hay que ser feminista para saberlo. Deberían cerrar las cuentas.	2
@AlbanoDante76: Tertuliana de @laSextaTV diciendo que la vida en la cárcel es "amable" con "siestas de dos horas". Cuanta miseria.	@AlbanoDante76 @laSextaTV Imagine que estoy en Alemania y me piden en mi tienda JUDÍO y una estrella de david.	2
@PatriotaNene: Argentinos que han venido a España y se permiten el lujo de querer trocear nuestra Nación, si quieres la deportación d...	@PatriotaNene Creo que la camisa tan apretada le impide un correcto riego sanguíneo al cerebro	2
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani no entendi	2
@fedevigevani: Domingo, nos volvemos a ver...	@fedevigevani mañana es lunes	2
@Soymarioruiz: Buenos días	@Soymarioruiz A esa hora yo andaba en el quinto sueño	2
@anabelenroy_tve: ¿Cuánto tenemos que aprender de ellos?... Qué delicia de video	@anabelenroy_tve Mi chihuahua es el ser más cariñoso del mundo	2
@anabelenroy_tve: ¿Cuánto tenemos que aprender de ellos?... Qué delicia de video	@anabelenroy_tve No quiero enfadar a nadie, pero no veo a ningún gato capaz de hacer algo parecido.	2
@EvilAFM: Estaba esperando este momento	@EvilAFM comprarse una Xbox one para disfrutar de la vida	2
@bryanmouque: Me encantaría poder dormir.	@bryanmouque @Coders_Boys hoy es el cumple de mi sobrina y sobri no que cumplen 4 años desde Guadalajara	2
@PabloMM: 5 tiparracos que violaron a una cría de 18 años en un portal piden respeto a su intimidad.	@PabloMM ahora eres juez??	2

Table A85: Examples of labeled data: Twitter (50)

@8choOficial: Termina esta frase:Hoy hay ____	@8choOficial EL HUMANO ES RETRASO Y NO TIENE CURA PA	2
@Rubiu5: Como joder la saga entera de Pokemon en 30 segundos	@Rubiu5 Alguien me explica que fue lo que dijo en inglés	2
@LSisco17: Sólo espero que la RAE acepte "Chavista" como sinónimo de mamagüevo.	@LSisco17 Es un acto amatorio no un insulto	2
@LSisco17: Sólo espero que la RAE acepte "Chavista" como sinónimo de mamagüevo.	@LSisco17 Se le dice directamente!!	2
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana.	@ecosdelbalon @XabiAlonso El trabajo tiene su recompensa.	2
@ecosdelbalon: Mirad con quién hemos pasado hoy la mañana.	@ecosdelbalon @XabiAlonso ¿Qué será lo próximo?	2
@tania_hermida: Ecuador pierde el rumbo. Michel Temer columnista invitado de nuestro diario público?@mfespinosaEC favor a tender	@tania_hermida @mfespinosaEC Con tanto éxito económico, ¿por qué tiene a duras penas un 2% de aceptación popular?...	2
@srtabebi: No me ha bajao.	@srtabebi Se acabó la bebida	2
@tolerancia0: Piñera en sobre las "empresas zombis" y el debate con Mónica Rincón y Daniel Matamala	@tolerancia0 @camandov Arriba los corazones... y las manos.	2
@ElJuanpaZurita: Bueno pues. Ahora sí. Noches noches noches noches.	@ElJuanpaZurita quiere irse	2
@itsandreix: El catalogo de ToyPlanet de juguetes si que sabe.	@itsandreix Precioso un poquito reconciliador con la raza humana	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio desagregado según sector socio económico?	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio acaso los pobres viven lo mismo que los ricos?	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio Eso significa vivir peor!!!	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio Cómo se llega a esa edad??	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio No es como lo mismo decir que el PIB per capita es el más elevado de latino América, no?	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio Agrégale a la curva el costo unitario del plan de salud...	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio entiendo que superamos por lejos en eficiencia a l resto	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio EEUU tiene una esperanza de vida muy baja comparada con cualquier país	2
@matiasdelrio: Este es el país que hemos construido y del que algunos sienten vergüenza.	@matiasdelrio como se nota que duermes vendado...	2
@Mel_Lezcano: Mis victimas de siempre jajaja @MoliniersReal @MicaViciante	@Mel_Lezcano @MoliniersReal @MicaViciante No se como agradecerelas.	2
@Juanmalpr: Eduardo Inda pillado hoy con Granados antes de la comparecencia en el Senado.¿Grandes amigos? á	@Juanmalpr ¿Con quién están, lo véis??	2
@AlvaroUribeVel: Petro creó 50 mil contratos burocráticos en Bogotá, quebró la ciudad y aspira a la Presidencia	@AlvaroUribeVel Es hora de que nosotros el pueblo esclavizado busquemos la manera de salvar nuestro país.	2
@Rubiu5: Doblaje Español confirmed	@Rubiu5 estoy en clase	2
@BangBangtan_Esp: En unas horas comienza "The late late show con James Corden" donde BTS se presentará.link para ver el show.	@BangBangtan_Esp @BTSxAMAs No quiero ir a la escuela	2
@BangBangtan_Esp: En unas horas comienza "The late late show con James Corden" donde BTS se presentará.link para ver el show.	@BangBangtan_Esp @BTSxAMAs unas dicen PM otras am cual creo	2
@Davanana: VAYANSE A LA VERGA	@Davanana @NiniIndigo1 Que es esto, la filosofía de la generación de gelatina?	2
@Davanana: VAYANSE A LA VERGA	@Davanana @ppmarino NINGUNA mujer le digan ERES HERMOSA la dañen.	2
@Davanana: VAYANSE A LA VERGA	@Davanana @ppmarino NO solamente hermosura EXTERIOR, Si no HERMOSA INTERIOR...	2
@acristofalo: Soy una buena alternativa para Benedetto, Sampaoli, teneme en cuenta.	@acristofalo no lo entendi	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @marcedeltiempo @FernandezAnibal En pocos meses será compañero de celda.	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @MariaLujan_Rey @FernandezAnibal Por que tiene celular ?	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @mis2centavos @FernandezAnibal Tiene internet el preso?	2
@JulioDeVido: Agradecemos la visita del compañero @FernandezAnibal a Marcos Paz.	@JulioDeVido @mis2centavos Dice @FernandezAnibal va a penetrar al arquitecto con este calor	2

초 록

인접 호응(linguistic alignment)은 대화에서 상대방이 직전에 말한 문장에 대해, 맥락에 맞는 말을 하는 것이다. 일반적으로 의사 소통에서, 화자는 상대방의 의견을 수락하거나 지지하기 위해 발화의 스타일을 변경하고, 상대방의 발언에 따라 의사 전달의 요지를 변경한다. 인접 호응 구조는 실생활에서 자주 발견되지만, 선행연구에 명확한 기준이 없었기에 먼저 어학적으로 이 문제를 정의하는 것에서 시작하였다. 인접 호응은 직전 발화의 맥락에 맞춰 대화를 이어가는 구조를 가리키며, 상대방의 앞선 발화를 긍정하거나 부정하는 경우, 그리고 앞에 나온 말을 반복하는 경우에 성립한다. 한편, non-alignment에는 앞에 나온 주제가 바뀌거나 대화의 흐름이 끊기고 더 이어지지 않거나 전혀 무관한 말을 하는 경우가 해당된다. 이 현상은 여러 상황에서 관찰되며 다양한 사회적 요인 및 원인에 따라 달라질 수 있다. 이것은 일종의 적응이나 수용 행동으로, 웹 상에서의 대화와 대면 상황에서 일반적으로 관찰된다. 분명 존재하나 실제로 규정하기 어려운 문제라는 점에서 전통적인 규칙기반 방식으로 인접 호응 구조를 분석하는 것은 쉽지 않다고 판단하였다. 따라서 그 해결책으로서 최신 딥러닝 방식을 활용하여 분석하고, 언어적 특성을 파악하고자 했다.

이 연구는 세 가지 신규성을 가지고 있다. 첫째, 인접 호응을 탐지하고 분류하기 위해 최초로 딥러닝을 적용하였다. 둘째, 최초로 스페인어에 인접 호응 개념을 적용하였으며, 이를 위해 스페인어 인접 호응 코퍼스와 인접 호응을 예측하는 모델을 만들었다. RNN과 CNN, Siamese network 모델을 설계하여 스페인 준구어 코퍼스에서 인접 호응 구문의 언어적 이론적 기준을 수립하고 분석하는 방법을 제안하였다. 셋째, 스페인어 어휘 표지자(marker)의 타당성을 검증하였다.

스페인어는 어휘 요소가 풍부하게 발달되어 있으며 대용어 등의 문장 성분이 명시적으로 다양하게 드러난 언어다. 우리는 어휘와 범주 특성에 따라 주요 표지

자를 추출하여 연구할 필요가 있다고 보고, 서울대학교 학부 대상 교과서 4권과 주제별 웹 트위터 데이터를 크롤링하여 수집하였다. 트위터 대화를 수집하기 위한 크롤러를 만들고, 수집한 스페인어 준구어 코퍼스에서 인접 호응이 발생하는 언어적 이론 기준을 정립하여 정답 데이터를 구축하였다. 특정 표지자에 의존하지 않는 모델을 위하여 문장 전체를 고려하는 딥러닝 방식을 적용하였다. 각각 RNN 과 CNN, Siamese network 기반의 인접 호응 여부를 예측하는 분류기를 만들어 적합한 표지자를 추출하고 검증하였다. 실험을 통해 일반적인 예상과 달리 특정 어휘 요소에 기반하여 스페인어 문장을 구분하는 것에는 무리가 있음을 확인하였다.

인접 호응 분류(alignment classification)를 학습하기 위해서 트위터 대화와 서울대학교 스페인어학습 교재 대화문을 사용하였다. 정립한 인접 호응 기준을 바탕으로 트위터 2,000쌍과 교재 1,384쌍 데이터를 분류하였다. 그리고 각 데이터를 1:1 비율로 증대(augmentation)한 8,400여 쌍(교재 alignment 1,960쌍, non-alignment 1,960쌍; 트위터 alignment 2,250쌍, non-alignment 2,250쌍) 데이터로 분류기(classifier)를 학습시켰다. 분류기의 성능은 이론에 기초하여 예상한 베이스라인(표지자가 답변 문장에 존재하면 인접 호응으로 분류)의 성능 35-55% 보다 높은 70-80%로 상당히 fair한 수준의 성능을 보였다.

학습한 인접 호응 분류기(alignment classifier)를 통해 총 50,000쌍(alignment 27,449쌍, non-alignment 22,551쌍)의 unlabeled 데이터를 분류하도록 하였고, 분류된 27,449쌍의 인접 호응 대화문을 이용하여 생성모델을 만들었다. 그 결과 labeled data만으로 생성을 하려 했던 것보다 양질의 인접 호응 문장 생성이 가능하였다. 이는 학습된 분류기를 사용하여, 데이터를 확보하는 방법을 통하여 사람의 노동력을 상당량 경감시키면서 학습에 필요한 데이터를 확보할 수 있음을 나타낸다. 우리는 대량의 분류된 스페인어 데이터를 기반으로 마커 검증을 하였고, 사람이 예상한 것 외의 마커를 발견하는 것이 가능했다. 이를 통해 언어학 연구에도 딥러닝 기술이 유용하게 사용될 수 있음을 확인하였고, 인접 호응 코퍼스 및 모델을 기반으로 사용자의 입력에 맞춰 맥락에 맞는 응답을 생성하는 시스템을 만들 수 있었다.

이 논문은 실제 언어 상황에서 발견되는 인접 호응 현상에 대한 이론적 개념을 정립하고, 쉽사리 규정하기 힘든 이 문제를 해결하기 위해 데이터를 수집하고 학습시켜 딥러닝을 통해 검증한 최초의 연구이다. 다양한 실험을 통해 인접 호응 구문을 탐지하고 모델을 만들어 표지자에 대한 검증을 시도하였으며, 인접 호응에 성립하는 답변을 생성하는 소기의 목적을 달성하였다. 지적인 문제는 현재 실무적으로도 어려우면서도 필요한 기술이며, 모델의 성능이 높지 않은 것은 문제 자체의 난이도를 반영하는 것으로 볼 수 있다. 데이터 부족 문제를 해결하고 사용자에게 적절하게 응답할 시스템을 개발하는 데에 필요한 기준 및 연구 방법으로 활용될 것으로 예상된다.

주요어 : 인접 호응, 스페인어 표지자, 인접 호응 분류, 표지자 검증, 인접 호응 답변 생성, 준구어 코퍼스

학번 : 2008-30759